THE BELL SYSTEM TECHNICAL JOURNAL

A JOURNAL DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

TABLE OF CONTENTS
AND
INDEX

VOLUME XV

1936

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XV, 1936

Table of Contents

JANUARY, 1936

Index to Volume XV

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

VOLUME XV JANUARY, 1936 NUMBER 1

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each. The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.

The Bell System Technical Journal

Vol. XV *January, 1936* *No. 1*

Long-Wave Radio Transmission Phenomena Associated with a Cessation of the Sun's Rays

By AUSTIN BAILEY and A. E. HARPER

The variations in long-wave radio field strength near the time of sunset on long transmission paths, which have been reported by many observers, were studied for the purpose of formulating rational methods of forecasting their time of occurrence. During some preliminary observations fair agreement was found between the time of minimum field and the sun's position relative to a particular point on the transatlantic path under observation. The more extended study of radio field variations during sunset periods and solar eclipses disclosed that in general no exact relationship could be established between the sun's position at any point and the occurrence of the minimum field.

Observations of field variations were made on radio signals at a number of different frequencies and over several paths. It was concluded that a characteristic sunset cycle of field variations is present on frequencies between 18 kc. and 68 kc. for transmission paths longer than 700 km. For paths less than 200 km. long, such variations are negligible. There is some evidence that the amplitude of these field variations is smaller at lower frequencies.

Analysis of the data presented indicates that long waves over long paths are transmitted predominately by "sky waves." From the data it was not possible to establish any satisfactory picture of the path followed. It was established, however, that empirical methods based on observations over a particular transmission path may be employed to forecast the approximate time of occurrence of field variations.

Introduction

COMMERCIAL transatlantic radio telephone service to Europe was inaugurated on January 7, 1927, using a long-wave circuit on a frequency of 60 kilocycles. The difficulties of maintaining a satisfactory long-wave circuit through that part of each day when sunset conditions existed along the transmission path had been recognized [1] * for several years prior to the opening of service. After short-wave facilities were added to the transatlantic service, it became important to coordinate operating times of the short-wave transmitters in relation to the long-wave channel to assure maximum reliability and efficiency of service. In order to do this, a more precise knowledge of the

* Numbers refer to appended list of references.

behavior of the long-wave circuit during the sunset period was needed, and it was for this purpose that many of the observations reported in this paper were made.

An analysis of these observations indicated that, aside from their practical application, some rather fundamental information concerning the probable mechanism of transmission on long waves could be obtained. To these observations other data were added, and all of this available material was systematically studied to determine the effect of the cessation of the sun's active rays on radio transmission at long wave-lengths. Although the results alone are rather inconclusive, they do provide sufficient evidence to indicate that the mechanism of long-wave transmission is in some ways the same as that of short-wave transmission and for the longer paths depends primarily on waves returned to the earth by layers in the atmosphere.

No attempt will be made in this paper to review the present status of radio transmission theory or of related geophysical phenomena, except in special cases where required to show the significance of our results. A background of the theory has been provided by many investigators, among whom Smith,[2] Pedersen,[3] Anderson,[4] Appleton,[5] Green,[6] Namba,[7] Yokoyama and Tanimura,[8] Hollingworth [9] and Heising [10] should be mentioned.

The analysis of the data taken during the present investigation indicated that the connection between solar and radio phenomena is effected through the agency of electromagnetic radiation, and not by means of low velocity corpuscular solar emission. A partial corroboration of this conclusion was secured by Schafer and Goodall,[11, 12, 13] the U. S. Bureau of Standards,[14] and others during several solar eclipses. The rather complete analysis of data taken during eclipses which has been made by Appleton and Chapman [15] also confirms this view.

Method of Measurement and Estimated Precision

The field strength of special 60-kc. test dashes transmitted from WNL, Rocky Point, New York, was measured at Houlton, Maine, Chatham, New Jersey, and Cupar, Scotland. The Houlton and Chatham measurements were made by means of meter comparisons with a calibrated local oscillator and it is believed that their relative accuracy is of the order of $\pm$ 0.1 db, although the absolute accuracy probably falls short of this figure by a considerable margin. Due to the comparatively slow rate of long-wave field variation, no effort was exerted to obtain better timing than $\pm$ 10 seconds. Since weak fields and high noise are common during the late afternoon hours on the transatlantic path, it is believed that the precision of the Cupar measurements necessarily falls short of that possible for local tests.

Measurements made at Houlton on telegraph traffic from Tuckerton, Nauen, Ongar and Rocky Point are believed to be subject to relative errors as great as or even greater than $\pm$ 2.5 db. Comparisons between a local oscillator and telegraph traffic ordinarily are made by means of a cathode ray tube, and errors are occasioned both by the difficulty of reading the tube scale and by the varying strength of telegraph signals whose intensity is a function of the keying speed probably because of sluggish antenna systems.

The measurements on special 60-kc. test signals are believed to be sufficiently precise to provide a satisfactory index of the phenomena. The measurements on telegraph traffic provide data for a qualitative estimate of the nature of the various phases of the phenomena but probably are of no great value in fixing the exact time of occurrence of field strength increments smaller than 2 db, which may represent the total variation of some phases of the diurnal cycle.

General Characteristics of Sunset Effect on Short Paths

The form of the average diurnal sunset cycle of received field strength, plotted as a function of the sun's angular altitude at some salient point on the great circle transmission path, is shown on Fig. 1 for several paths. With the exception of the shortest path of 122 km. between Rocky Point and Chatham, New Jersey, a well defined typical sunset cycle was observed in all cases. If we assume that the sunset dip is due directly or indirectly to the cessation of the active solar rays in the upper atmosphere, the absence of characteristic large field variations at sunset for short paths may be ascribed to a predominance of low elevation transmission which does not enter the layers ionized by the sun. If rectilinear this transmission would pass ½ km. above Chatham on the 122-km. path, due to tangency with the earth's surface. The possibility of transmission between two points 122 km. apart on the earth's surface without a "sky wave," [5] therefore requires that the ray be bent around the spherical contour of the earth by diffraction and atmospheric refraction.[3, 16]

The daily occurrence of a sunset dip on paths of 700 km. and longer may be explained in several ways, all of which, however, require the existence of a downcoming ray. For example, the sunset phenomena may be due to a "fault" in the reflecting ionized layer,[7] or to interference between two reflected rays or between a reflected ray and a ground wave.[9, 17, 18]

Briefly, in accordance with the "fault" hypothesis, the shadow cast upon the ionized reflecting layer by the sun's active rays tangent to the earth's surface, or to an opaque atmospheric layer concentric with the earth which hereafter will be called the occulting layer, produces a

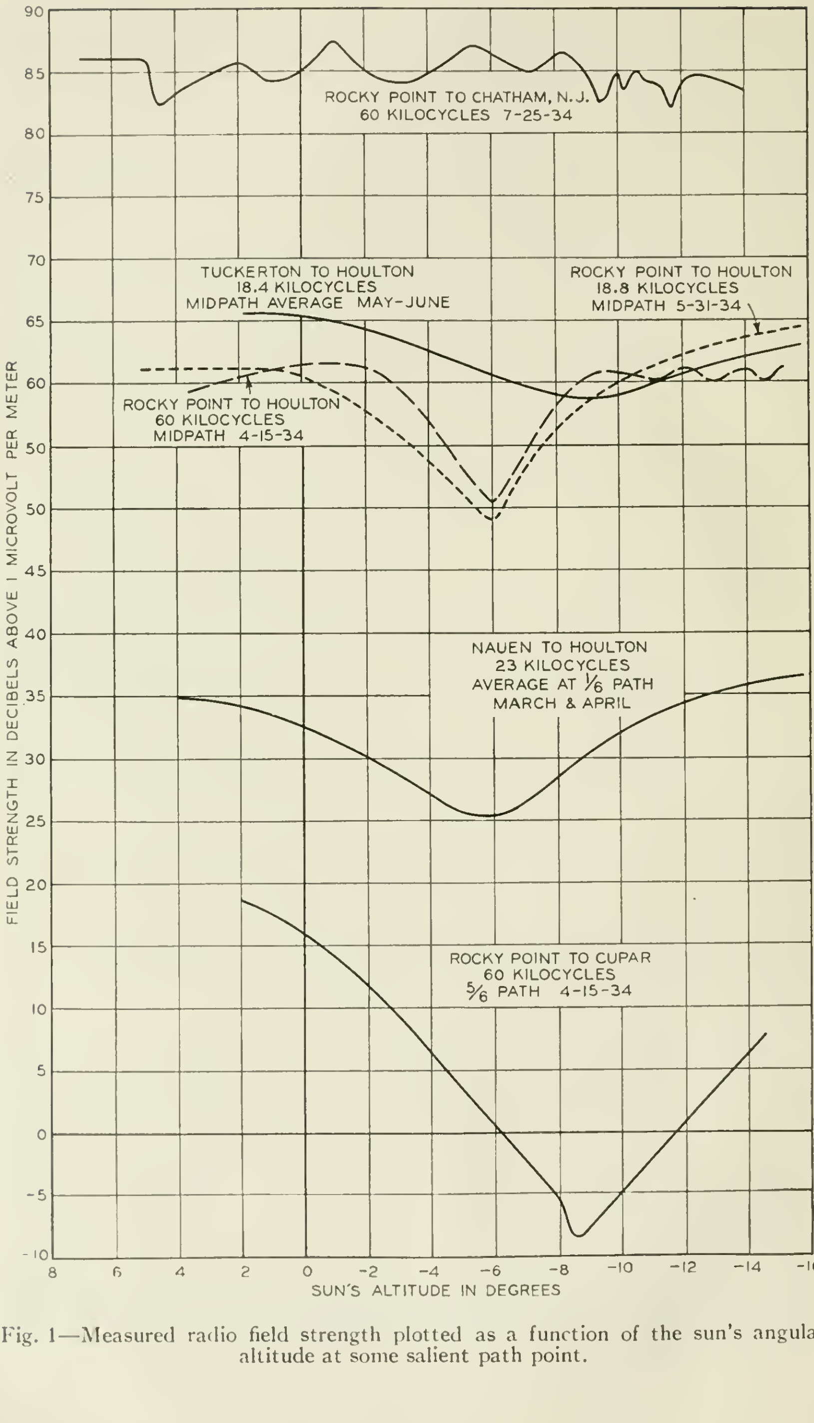


Fig. 1—Measured radio field strength plotted as a function of the sun's angular altitude at some salient path point.

"fault" in the otherwise uniform daylight ionized layer surface. The form of fault presumably is a ring generated by the intersection of the shadow cone of the occulting layer with the spherical reflecting layer surface. As the fault passes over the reflected radio ray apex, reception of the optimum ray may be impeded, either because of scattering due to an irregular reflecting surface,[3] or on account of the passage of the ionization density through a value defined by the Brewster angle, thereby changing the mode of transmission from daylight metallic reflection to night-time refraction.[7]

One of the objects of this analysis was to find the angular altitudes of the sun at some fixed point on the transmission path, coinciding with the phases of the sunset cycle of field variation at the receiving station. It was expected that with these data available over the period of a year, it would not be difficult to locate the salient point on the transmission path affected by the cessation of the sun's rays by means of Sumner lines of position [19,20] provided, of course, that the physical structure of the layers showed negligible variation. From the above-mentioned data it would also be possible to compute the distance between the occulting and reflecting layers [21] and to accurately predict the time of occurrence of the sunset minimum. As shown in Tables I–IV, annual and fortuitous variations in the time of occurrence of the phases of the phenomenon prevent the satisfactory application of this method to long path effects, and when applied to the Rocky Point-Houlton path the point so located apparently is situated a considerable distance to the southwest of the most southerly path terminal. This presumably indicates that, if the phenomena take place at a fixed location on the transmission path, there must either be an annual variation of considerable magnitude in the effective distance between the reflecting and occulting layers, or the mechanism involved must be considerably more complex than that initially assumed.

Frequency Range of Phenomena

The characteristic diurnal sunset cycle was found on all frequencies studied during this investigation, the scope of which was limited to frequencies between 18 and 68 kilocycles. Available published data taken by other investigators indicate that the pronounced sunset minimum characteristic of long-wave transmission is not observed at broadcast frequencies.[6,22] This may be due to the failure of transmission to improve after the sunrise drop, giving an all day minimum, or, if the minimum is due principally to interference bands, the fineness of band pattern at high frequencies may prevent the phenomenon from being recognized.

Observations on GBR at 16 kilocycles at Houlton during the latter part of April, 1934, showed little evidence of a minimum. These measurements were made on telegraph traffic and if a minimum occurred its amplitude must have been less than the observational errors of this method of measurement. These results seem to be confirmed by the observations of Yokoyama and Tanimura [8] which show a pronounced decrease in the amplitude of the sunset cycle at frequencies below 17 kilocycles, while frequencies slightly above this value display the characteristic effects. This apparent difference in the behavior of 18 and 16-kc. transmission, seems rather unusual and if real may have some geophysical significance.

Results

An examination of the data taken on the Rocky Point-Houlton 60-kc. transmission path as plotted on Figs. 2 and 3 discloses that of the ten cases plotted, nine show that a minimum in measured field occurs 22–26 minutes or an average of 23 minutes after surface sunset at Rocky Point, corresponding to an altitude of the sun of 4 to 5½ degrees below the horizon, and in a single case at 18 minutes after sunset on August 29, 1934. The results in the winter months seem less regular than in summer and the cases of 2/25/34 and 1/21/34 are especially noteworthy in that they show multiple minima, the 23-minute minimum being subsidiary to a minimum occurring about an hour after Rocky Point sunset on 1/21/34. The fact that minima occur at a nearly constant interval after Rocky Point surface sunset, which occurs at the same time that the sun's rays are tangent to concentric elevated layers above this point, rather than sunset at some other path point, is believed to be the result of a fortuitous combination of circumstances rather than a rational relationship between these times. This hypothesis is strengthened by data shown in Table I taken over the same path at 18 kilocycles and data in Table II taken at 18 kilocycles over the 900-km. Tuckerton-Houlton path which do not show this same constant relationship with the time of sunset at the transmitting station. Of three observations obtained over the Rocky Point-Houlton 18-kc. transmission path the minimum in field occurred 34 minutes after sunset at Rocky Point in two cases but on the other day it occurred 22 minutes after sunset at Rocky Point. On all three occasions, however, the minimum was between 35 and 37 minutes ($-6°$ to $-6.5°$ altitude of sun) after surface sunset at the mid-path point. As in the case of the 60-kc. path for which data are shown on Figs. 2 and 3, the field began to fall at mid-path surface sunset.

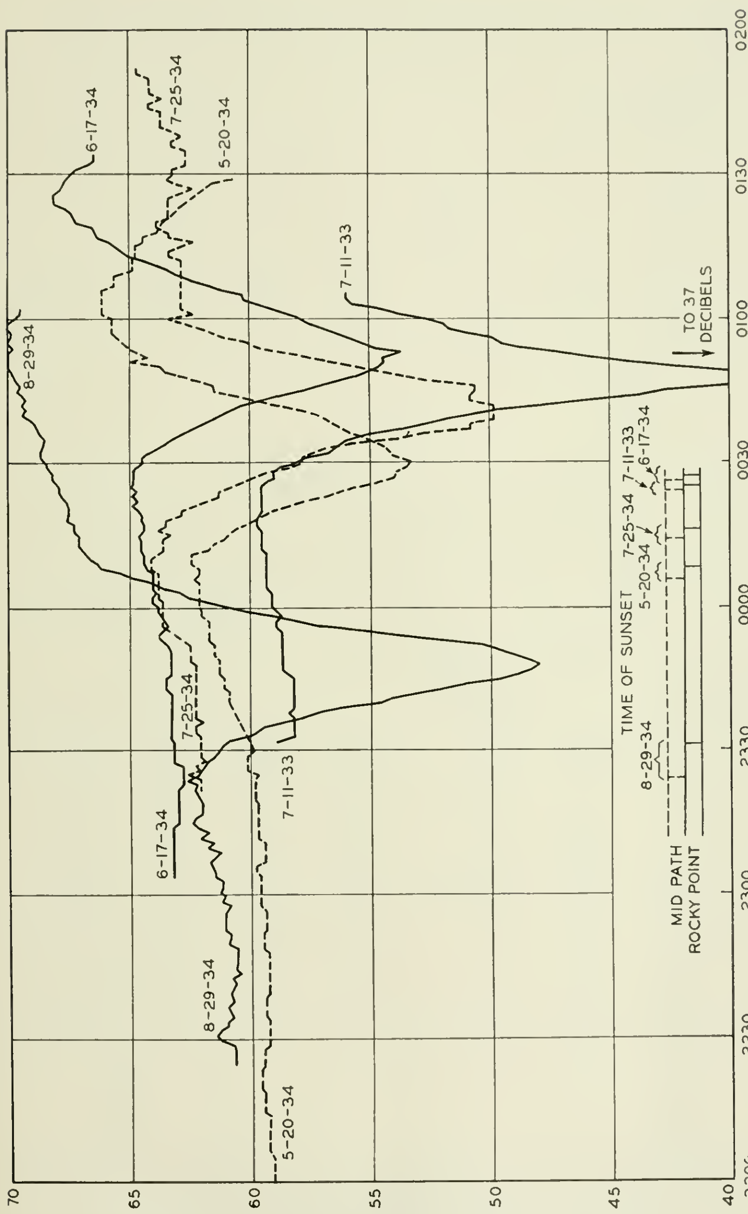


Fig. 2—Field strength variations on the Rocky Point-Houlton 60 kc. transmission path during the summer months.

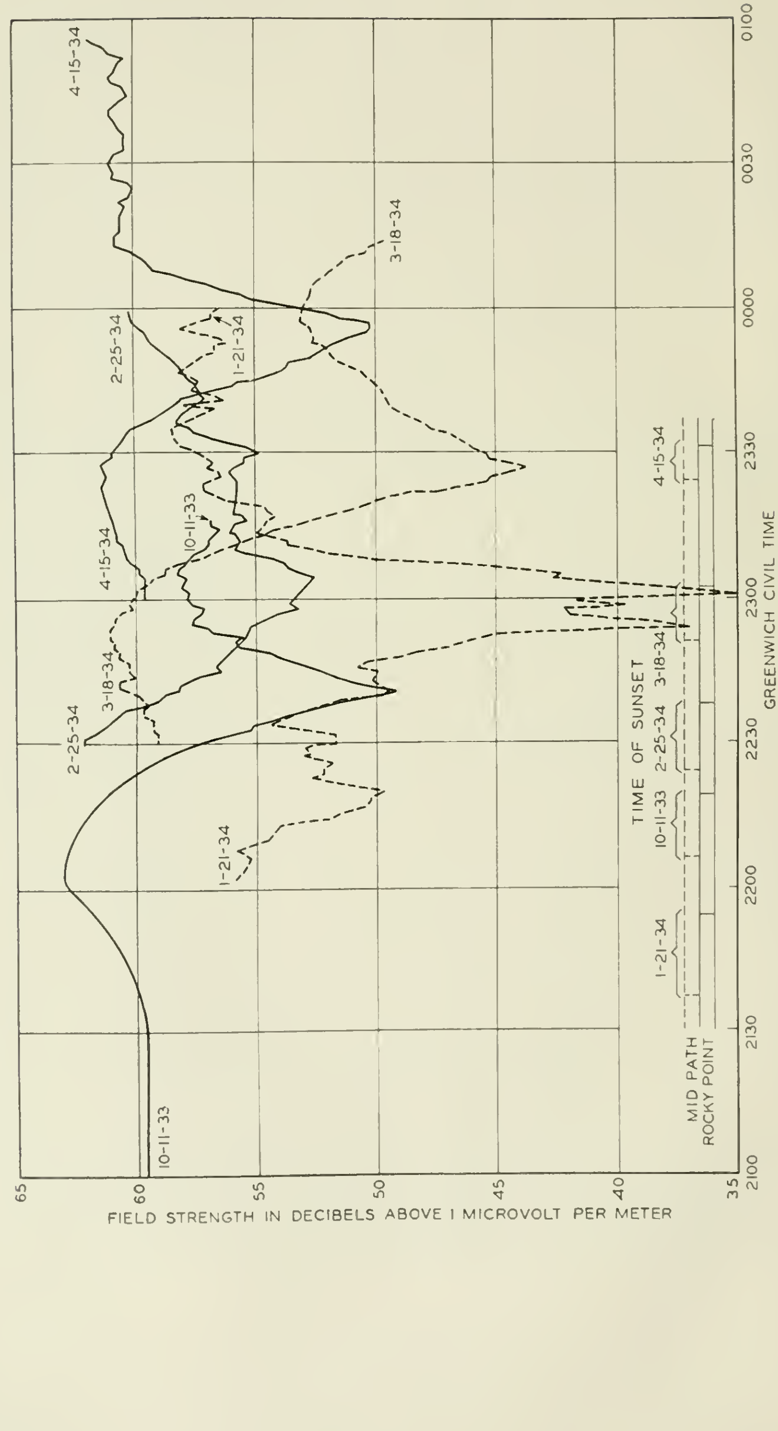


Fig. 3—Field strength variations on the Rocky Point-Houlton 60 kc. transmission path during the winter months.

TABLE I

ROCKY POINT-HOULTON—18.8 KC.

Distance 700 Km. Average Minimum 13 Db Below Field Immediately Preceding Drop

Date	Time of Sunset, GCT		Minimum Field		Interval Between Sunset and Minimum		
	Rocky Point	Houlton	Time GCT	Sun's Altitude at Midpath Degrees	Mid-path	Rocky Point	Houlton
6/ 1/34....	0016	0014	0050	−6.0	0:35	0:34	0:36
8/ 2/34....	0009	0003	0043	−6.7	0:37	0:34	0:40
10/21/34....	2204	2137	2226	−7.4	0:35	0:22	0:49
			Average	−6.7	0:36	0:30	0:42

TABLE II

TUCKERTON-HOULTON—18.4 KC.

Distance 900 Km. Average Minimum 7.9 Db Below Field Immediately Preceding Drop

Date	Time of Sunset, GCT		Minimum Field		Interval Between Sunset and Mininum		
	Tuckerton	Houlton	Time GCT	Sun's Altitude at Midpath Degrees	Mid-path	Tuckerton	Houlton
4/28–29/34 .	2349	2335	0030	− 9.0	0:48	0:41	0:55
4/30–5/1/34.	2351	2337	0033	− 9.0	0:49	0:42	0:56
5/ 1– 2/34 .	2352	2338	0040	−10.0	0:55	0:48	1:02
5/ 3– 4/34 .	2354	2342	0045	− 9.6	0:57	0:51	1:03
5/ 9–10/34 .	2359	2349	0053	−10.2	1:00	0:54	1:04
5/10–11/34 .	0000	2350	0047	− 9.4	0:52	0:47	0:57
5/11–12/34 .	0001	2351	0027	− 5.9	0:31	0:26	0:36
5/17–18/34 .	0006	2359	0048	− 8.0	0:46	0:42	0:49
5/21–22/34 .	0010	0004	0046	− 6.9	0:39	0:36	0:42
5/24–25/34 .	0013	0007	0050	− 7.0	0:40	0:37	0:43
5/28–29/34 .	0016	0011	0115	− 9.8	1:02	0:59	1:04
6/ 4– 5/34 .	0021	0017	0117	− 9.1	0:58	0:56	1:00
6/ 7– 8/34 .	0023	0019	0115	−11.0	0:54	0:52	0:56
6/11–12/34 .	0025	0022	0120	− 9.0	0:57	0:55	0:58
6/14–15/34 .	0027	0023	0122	− 9.3	0:57	0:55	0:59
			Average	− 9.5	0:51	0:47	0:55

The 18-kc. Tuckerton-Houlton 900-km. path of nearly the same azimuth as the Rocky Point-Houlton path, showed a minimum occurring irregularly at from 31 to 62 minutes, or an average of 51 minutes after mid-path sunset, corresponding to an average sun's altitude of − 9.5°. Since the data of Table II are subject to the errors inherent

to measurements made on telegraph traffic, a portion of the large variation in time may be due to experimental errors.

The wave-like variations in field intensity observed on the Rocky Point-Houlton 60-kc. path on 1/21/34 and 4/15/34 have the appearance of interference fringes,[9, 17] and if explained on this basis the question at once arises whether or not interference is the principal cause of the sunset cycle. If daylight communication is accomplished either by means of a single reflected ray in combination with a ground wave, or by the resultant of a number of reflected rays of nearly constant complex propagation constants, at sunset the occultation of the ionizing rays from the transmission medium by tangency with an occulting layer, might reasonably be expected to produce variations in both the real and imaginary portions of the propagation constant of the medium, thereby causing interference fringes through the variation of the relative phase of different rays.

As an example, we may assume that on the Rocky Point-Houlton transmission path the ground ray provides the principal agency of daylight communication by means of atmospheric refraction and diffraction. At the approach of sunset, reduced ionization between the earth and the reflecting layer reduces the attenuation in the path of the reflected ray, producing a resultant received field which is a function of the relative intensity and phase of the two waves. As the effect of the sun's active rays becomes still less, the decreased ionization of the layers produces variations in the phase of arrival of the reflected wave, either through changes in the propagation constant, or on account of greater path length occasioned by an increased virtual height of the reflecting layer.

Now since the resultant received field is the vector sum of the two rays, when one ray is much smaller than the other, variations in their relative phase will produce small amplitude fluctuations in the resultant, with maxima and minima equal to the sum and difference of the two components. As the two components approach equality, however, the maxima will approach twice the intensity of one component ray, while the minima will approach zero, thereby producing a very deep "dip" in the received field.

The "D" Layer

It has been suggested by Heising,[10] Appleton [23] and Chapman [21] that passage through a low-elevation layer of ozone produced by solar ionization, is one of the principal causes for the daylight attenuation of the reflected ray.[2] Radio transmission measurements at broadcast frequencies, reported by the U. S. Bureau of Standards [22] and by the Australian Council for Scientific and Industrial Research,[6] show a rapid

decrease in sky-wave attenuation beginning shortly before tangential sunset and continuing for from one-half to two hours after sunset. On the 60-kc. long-distance path the evidence of the initial phases of this phenomenon is not so well defined, due probably to the gradual nature of the change, and to the characteristic sunset minimum which occurs shortly after sunset. Recent measurements made on the transatlantic radio telephone channels, however, indicate a presunset rise in field of about 2 db, beginning when the sun's altitude becomes low enough to cause an appreciable lengthening of the atmospheric ray path, thereby decreasing the intensity of the active solar rays. A presunset increase is very apparent on the Rocky Point-Houlton 60-kc. measurements, beginning at about 40 minutes before, and continuing to a maximum at the instant of mid-path sunset. There is a possibility, however, that this phase may be due to interference phenomena.

More conclusive evidence of the effect of a reduction in the intensity of the sun's ionizing rays is provided by long-wave field strength measurements made during the solar eclipses of January 24, 1925, and August 31, 1932. The data taken in 1925 were secured by means of an automatic recorder on the Rocky Point-Belfast 57-kc. path and manual measurements were made at the European receiving stations. In 1932, automatic recorders were used on both the Rocky Point-Houlton 60-kc. path and the Rugby-Houlton 68-kc. path. In all cases where automatic recorders were used the data were abstracted from the record and replotted for reproduction.

Figures 4 and 5 show the eclipse circumstances and the concurrent variations in the measured radio field strength. The Chedzoy, New Southgate, and somewhat less clearly the Aberdeen measurements in 1925, show the results of reduced attenuation almost immediately after the first darkening of the transmission path. For the 1932 data this effect is clearly evident on the Rocky Point-Houlton measurements.

The sudden increase in field to be expected at the beginning of the eclipse at Belfast in 1925 is not apparent from the data, and its absence may be due either to improper recorder operation or to some fortuitous phenomenon peculiar to that particular eclipse, such as, for example, the possibility that the phases of ground and sky waves were in quadrature. The Rugby-Houlton observations in 1932 likewise are ambiguous because the true eclipse effect is complicated by superposed sunset effects originating at the eastern path terminal, and the observed increase in field may be due to these rather than to the eclipse.

In all cases except the two mentioned above, the increase in field was followed by a rapid drop, with a minimum occurring at the approximate time the totality shadow crossed the transmission path.

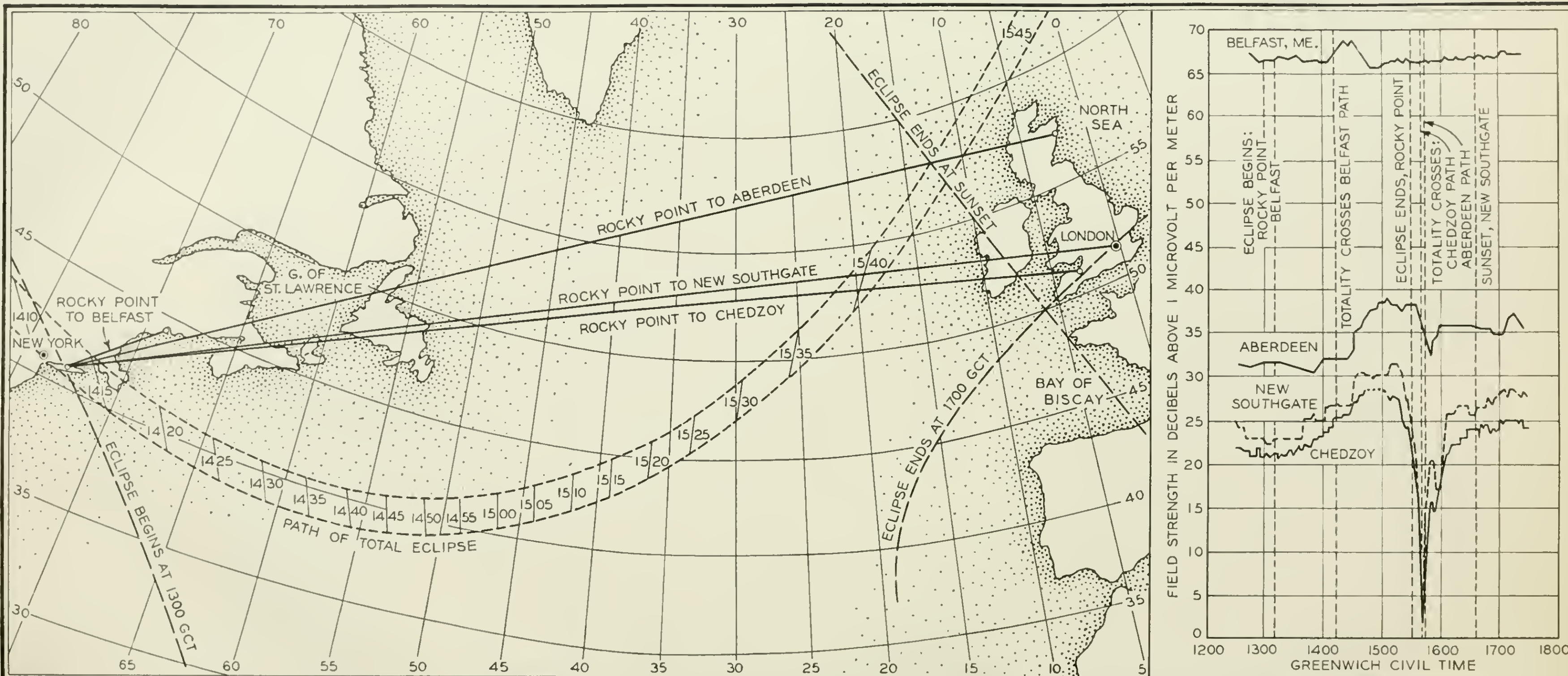


Fig. 4—Relation between 57 kc. radio transmission and circumstances of the solar eclipse of January 25, 1925.

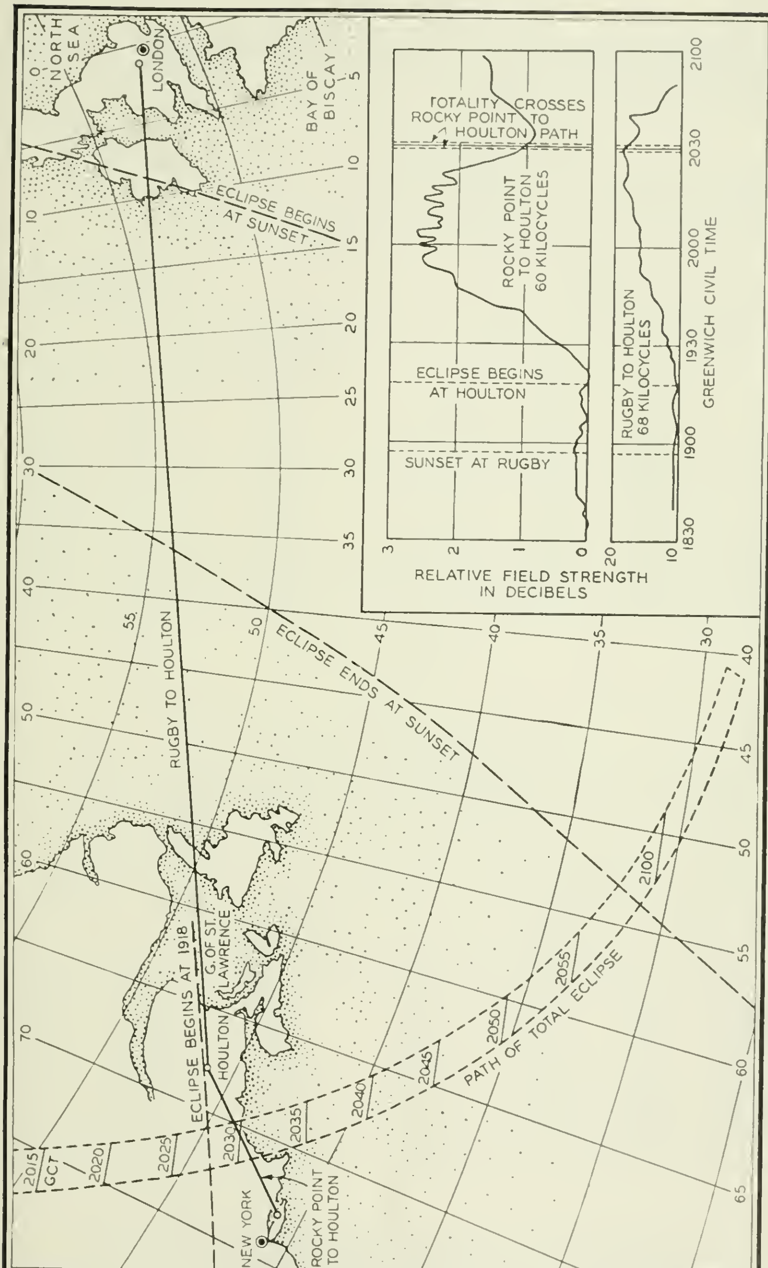


Fig. 5—Relation between 60 kc. radio transmission and circumstances of solar eclipse of August 31, 1932.

Long Transmission Paths

An examination of long-distance data of Tables III and IV discloses the following facts regarding these phenomena.

TABLE III

Nauen-Houlton—23 Kc.

Distance 5600 Km. Average Minimum 13 Db Below Field Immediately Preceding Drop

Date	Time of Sunset, GCT		Minimum Field		Interval Between Sunset and Mininum	
	Nauen	1/6 Path Point	Time GCT	Sun's Altitude at 1/6 Path Point Degree	Nauen	1/6 Path Point
3/12/34.....	1706	1756	1838	−6.2	1:32	0:42
3/19/34.....	1718	1810	1836	−4.2	1:18	0:26
3/22/34.....	1723	1816	1900	−5.5	1:37	0:44
3/26/34.....	1730	1824	1850	−4.2	1:20	0:26
3/29/34.....	1735	1830	1903	−5.5	1:28	0:33
4/ 2/34.....	1742	1839	1925	−6.7	1:43	0:46
4/ 5/34.....	1747	1845	1857	−2.5	1:10	0:12
4/ 9/34.....	1754	1853	1940	−7.0	1:46	0:47
4/10/34.....	1756	1855	1944	−7.0	1:48	0:49
4/12/34.....	1759	1859	1950	−7.2	1:51	0:51
			Average	−5.6	1:33	0:37

TABLE IV

Rocky Point-Cupar—60 Kc.

Distance 5172 Km. Average Minimum 25 Db Below Field Immediately Preceding Drop

Date	Time of Sunset, GCT		Minimum Field		Interval Between Sunset and Minimum	
	Cupar	5/6 Path Point	Time GCT	Sun's Altitude at 5/6 Path Point Degrees	Cupar	5/6 Path Point
10/11/33....	1724	1817	1910	− 8.0	1:46	0:53
11/22/33....	1554	1644	1807	−10.4	2:13	1:23
12/20/33....	1548	1626	1756	−10.6	2:08	1:30
2/25/34....	1736	1830	1958	−12.5	2:22	1:28
3/18/34....	1820	1916	2029	−10.7	2:09	1:13
4/15/34....	1917	2015	2117	− 8.4	2:00	1:02
6/17/34....	2101	2206	2322	− 6.6	2:21	1:16
			Average	− 9.6	2:08	1:15

Rocky Point-Cupar 60-Kc. Path, Table IV

1. Minimum field does not follow surface sunset at any point on the transmission path by a constant time interval independent of season.

2. The altitude of the sun as measured from the most easterly path apex of an hypothetical three-reflection path varies from approximately $-6°$ in summer to $-13°$ in winter at the instant of minimum field, and computation shows that there is no point on the path at which the altitude remains constant.

3. Daily irregularities of considerable magnitude are often noted.

Ongar-Houlton Path

1. Multiple minima of irregular distribution seem characteristic of this path.

2. Averages of field strength as a function of the sun's altitude at three path apices show separate minima when the sun is at about $-6°$ during May and June.

Nauen-Houlton, Table III

1. The time of minimum field varies as much as 40 minutes within a few days.

2. An average of the March-April data indicates that the minimum takes place when the sun's altitude is approximately $-5.6°$ at the first of three path apices, or 37 minutes after sunset at the earth's surface under this point.

Conclusions

Based solely upon the rather meagre data gathered during this study, and therefore subject to further confirmation before being considered generally applicable to all long-wave transmission paths, our conclusions may be recapitulated as follows:

1. A characteristic cycle of events accompanies the cessation of the sun's active rays, consisting of an initial increase in field during the reduction of solar intensity, followed by a minimum received field after the sun's rays are cut off at the earth's surface. This cycle, occurring both at sunset and during total solar eclipses, is typical of long-wave transmission in the 18–70-kc. band over paths in excess of a few hundred kilometers.

2. The time interval between sunset at any point on the transmission path and the instant of minimum field has an annual and an apparently fortuitous daily variation. As a result of these variations the minimum does not occur when the sun is at a fixed angular altitude at any point on the transmission path.

3. The time interval between sunset at some fractional path point and the instant of minimum field, and likewise the angular altitude of the sun referred to the plane of the horizon at such a point, increases with both the path length and the wave-length, and is maximum during the winter months.

4. The amplitude of variation of the sunset cycle apparently is reduced greatly at frequencies below 17 kilocycles and on paths shorter than 200 km.

5. Evidence of interference fringes on some of the observations suggests the possibility that the main sunset minimum may be the result of interference phenomena.

6. The fact that the phases of the radio transmission cycle closely follow the optical eclipse circumstances indicates that the radio phenomena must be related to solar radiation of velocity similar to that of light.

7. On transatlantic paths during the spring and summer months the average sunset minimum occurred when the sun was approximately 6° below the horizon at one-sixth of the total path length from the eastern terminal. Since these paths varied considerably in latitude and length, the phenomena may be related to effects occurring at the most easterly apex of a three-reflection path. The data obtained in this investigation, however, are not sufficient to definitely establish the generality of this hypothesis.

8. Empirical rules may be formulated for the prediction of the time of occurrence of various phases of the sunset cycle on short transmission paths. For example, the beginning of the drop in field on the 60-kc. Rocky Point-Houlton path occurs at mid-path surface sunset, and the minimum occurs at approximately 23 minutes after surface sunset at Rocky Point. On longer paths larger fortuitous variations occur and available data fail to connect the time of the minimum with sunset at any point on the transmission path. Representative curves drawn from the data, although subject to random errors, provide an empirical method for the prediction of the approximate time of occurrence of the phenomena and are of service in traffic and power scheduling.

Acknowledgments

The authors are indebted to the engineers of the British Post Office and the operating personnel of the American Telephone and Telegraph Company, for their generous cooperation in measurement schedules, and to their colleagues in Bell Telephone Laboratories for their assistance and constructive criticism in the preparation of this paper.

References

1. "Transatlantic Radio Telephony," H. D. Arnold and Lloyd Espenschied, *Jour. A.I.E.E.*, August 1923; *Bell Sys. Tech. Jour.*, October 1923.
2. 24th Kelvin Lecture, "The Travel of Wireless Waves," by Sir Frank E. Smith, *Inst. E.E.*, Vol. 9, No. 25, March 1934.
3. "The Propagation of Radio Waves," by P. O. Pedersen.
4. "Transatlantic Radio Transmission and Solar Activity," C. N. Anderson, *Proc. I.R.E.*, Vol. 16, pp. 297–347, March 1928.

5. "On Some Direct Evidence for Downward Atmospheric Reflection of Electric Waves," by E. V. Appleton and M. A. F. Barnett, *Proc. Royal Soc.*, 109, 1925, pp. 621–641.
6. "A Preliminary Investigation of Fading in New South Wales," by A. L. Green and W. G. Baker, Bulletin No. 63, Radio Research Board Report No. 4, Council for Scientific and Industrial Research, Commonwealth of Australia, 1932.
7. "General Theory of the Propagation of Radio Waves in the Ionized Layer of the Upper Atmosphere," Shogo Namba, *Proc. I.R.E.*, February 1933.
8. "Some Long Distance Transmission Phenomena of Low-Frequency Waves," I. Yokoyama and I. Tanimura, *Proc. I.R.E.*, February 1933.
9. "The Propagation of Radio Waves," by J. Hollingworth, *Jour. I.E.E.*, Vol. 64, 1925–1926, pp. 579–589.
10. "Experiments and Observations Concerning the Ionized Regions of the Atmosphere," by R. A. Heising, *Proc. I.R.E.*, Vol. 16, January 1928, pp. 75–99.
11. "The Effect of the Recent Solar Eclipse on the Ionized Layers of the Upper Atmosphere," Schafer and Goodall, *Science*, Vol. 76, November 11, 1932, pp. 444–446.
12. "Radio Studies of the Ionosphere," by Schafer and Goodall, *Nature*, September 30, 1933, pp. 521–522.
13. "Ionosphere Measurements During the Partial Eclipse of the Sun of February 3, 1935," by Schafer and Goodall, *Nature*, March 9, 1935, pp. 393–394.
14. "Radio Observations of the Bureau of Standards During the Eclipse of August 31, 1932," by S. S. Kirby, L. V. Berkner, T. R. Gilliland and K. A. Norton, *Proc. I.R.E.*, Vol. 22, No. 2, February 1934.
15. "Report on Ionization Changes During a Solar Eclipse," by E. V. Appleton and S. Chapman, *Proc. I.R.E.*, Vol. 23, June 1935, pp. 658–669.
16. "Radio Propagation Over Spherical Earth," by C. R. Burrows, *Proc. I.R.E.*, Vol. 23, May 1935, pp. 470–480.
17. "Summary of Progress in the Study of Radio Wave Propagation Phenomena," by G. W. Kenrick and G. W. Pickard, *Proc. I.R.E.*, Vol. 18, April 1930, pp. 649–668.
18. "Phase Interference Phenomena in Low-Frequency Radio Transmission," by G. W. Kenrick and G. W. Pickard, *Proc. I.R.E.*, Vol. 22, March 1934, pp. 344–358.
19. "American Practical Navigator," Bowditch, *Hydrographic Office Publication H.O.* No. 9, Government Printing Office, Washington, D. C.
20. "The Sumner Line of Position," *U. S. Hydrographic Office Publication H.O.* No. 203, Government Printing Office, Washington, D. C.
21. Bakerian Lecture, "Some Phenomena of the Upper Atmosphere," by S. Chapman, *Proc. Royal Soc.*, Vol. 132, No. A820, August 1, 1931.
22. "An Analysis of Continuous Records of Field Intensity at Broadcast Frequencies," by K. A. Norton, S. S. Kirby and G. H. Lesler, Research Paper RP752, *Jour. of Research, Nat. Bureau of Standards*, Vol. 13, December 1934.
23. "Some Measurements of the Equivalent Height of the Atmospheric Layer," by E. V. Appleton, *Proc. Royal Soc.*, Vol. 126, 1929–1930, p. 543.

APPENDIX I

Position of Intermediate Points on the Transmission Path

The method of determining the transmission path parameters and the position of intermediate path points spaced a given distance from a path terminal is given in detail below.

Let φ = Latitude,

L_o = Longitude,

D = Distance, nautical miles between subscript points,

C = Path azimuth at subscript point,

Subscript a denotes sending station,

Subscript b denotes receiving station,
Subscript x denotes intermediate point,
$L_{o_{ab}}$ = Difference in longitude between a and b.

By the law of cosines, in a spherical triangle of sides abc and angles ABC, if we are given two sides and the included angle, the side opposite the given angle may be computed from (1) below.

$$\cos a = \cos b \cos c + \sin b \sin c \cos A \tag{1}$$

or substituting geographical coordinates

$$\cos D_{ab} = \sin \varphi_a \sin \varphi_b + \cos \varphi_a \cos \varphi_b \cos L_{o_{ab}}. \tag{2}$$

This may be made more convenient for logarithmic computation by writing it in the following form:

$$\text{hav } D_{ab} = \text{hav } (\varphi_a - \varphi_b) + \cos \varphi_a \cos \varphi_b \text{ hav } L_{o_{ab}}. \tag{3}$$

Now by the law of sines

$$\begin{cases} \sin C_a = \dfrac{\cos \varphi_b \sin L_{o_{ab}}}{\sin D_{ab}} = \cos \varphi_b \sin L_{o_{ab}} \csc D_{ab}, \\[2ex] \sin C_b = \dfrac{\cos \varphi_a \sin L_{o_{ab}}}{\sin D_{ab}} = \cos \varphi_a \sin L_{o_{ab}} \csc D_{ab}. \end{cases} \tag{4}$$

Equations (3) and (4) above determine the great circle distance between "a" and "b," the azimuth of "a" from "b," and "b" from "a." To find the position of a point "x" located a fraction of the total distance between "a" and "b" we again substitute in (1), obtaining

$$\begin{cases} \sin \varphi_x = \sin \varphi_b \cos D_{xb} + \cos \varphi_b \sin D_{xb} \cos C_b, \\ \sin \varphi_x = \sin \varphi_a \cos D_{xa} + \cos \varphi_a \sin D_{xa} \cos C_a, \end{cases} \tag{5}$$

and by the law of sines

$$\begin{cases} \sin L_{o_{ax}} = \dfrac{\sin C_a \sin D_{xa}}{\cos \varphi_x} = \sin C_a \sin D_{xa} \sec \varphi_x, \\[2ex] \sin L_{o_{bx}} = \dfrac{\sin C_b \sin D_{xb}}{\cos \varphi_x} = \sin C_b \sin D_{xb} \sec \varphi_x. \end{cases} \tag{6}$$

The latitude and longitude of intermediate point "x" are therefore determined by equations (5) and (6) above.

APPENDIX II

Determination of Sun's Altitude and Azimuth

The angle of the sun to the horizon and to the meridian plane at any hour may be computed by methods similar to the above. For this case we have a celestial triangle whose sides are a meridian through the observer's zenith, a meridian through the sun, and a great circle through the sun and the zenith. The arc subtended by the pole and zenith is the complement of the observer's latitude "φ," the arc subtended by the pole and the sun is the complement of the sun's declination "d" (celestial latitude), and the angle "t" at the pole between these two arcs is the sun's hour angle. With these two sides and included angle we may compute the arc between the sun and the zenith (complement of the altitude "h") and the sun's azimuth "z" which is the angle between the meridian containing the zenith and the great circle passing through the zenith and the sun.

By the law of cosines (1) above

$$\sin h = \sin d \sin \varphi + \cos d \cos \varphi \cos t \tag{8}$$

and by the law of sines

$$\sin Z = \frac{\sin t \cos d}{\cos h}. \tag{9}$$

Values of h and z as a function of φ, d and t are tabulated in convenient form in *hydrographic office publication H.O.* No. 203. The sun's declination and the computed times of sunset may be obtained from the *American Nautical Almanac.*

The Corrosion of Metals—I. Mechanism of Corrosion Processes

By R. M. BURNS

This paper outlines the application of electrochemical methods to corrosion investigations. It discusses the position of the potential of a metal against its environment and the trend of this potential with time, pointing out that it is thereby possible to determine whether the corrosion process is controlled by reactions occurring at the anodic areas, the cathodic areas, or both; that is, whether there is a tendency toward passivity, inhibition or progressive attack. Measurements of film stability whether in terms of the leakage current which may be passed through the film or in terms of the amount of film forming material required to produce passivity or the amount of film destroying material required to render a metal active, furnish information as to the quality of corrosion resistant films. Measurements of the rate at which a film forms on a metal when placed in a film-forming environment throws light on its relative surface reactivity, and such information is of assistance in determining the rate of corrosion in homogeneous corrosive environments or the rate of passivation in the film-forming environments. On the basis of such measurements and with a chemical knowledge of the environments in which metals are used as well as the composition and physical state or structure of the metals, it is possible to predict corrosion behavior and to obtain an understanding of corrosion problems usually not possible by ordinary empirical corrosion tests.

ALL metals are corrodible under the appropriate circumstances. The most important metal industrially, iron, is probably the most corrodible under ordinary conditions. Many estimates have been made of the value of iron and steel products destroyed by corrosion.[1] While much depends upon the basis of calculation it seems reasonable to conclude that the annual cost of corrosion in this country is of the same order as the interest on the public debt or nearly one third of the cost of the federal government in normal times. The common non-ferrous metals—zinc, lead, copper, aluminum, nickel and tin—are more resistant to corrosion largely because of their tendencies to form protective surface films. In the atmosphere under favorable circumstances tests have indicated, for example, that in the form of sheet 0.03 inch in thickness and exposed on one side as in the case of roofings, zinc, copper and lead if mechanically undisturbed would resist corrosion for more than one, two and three centuries respectively.[2] Once a protective film is formed it may preserve the metal indefinitely. Under other circumstances these metals may readily corrode. Contact with large inert soil particles may result in the perforation of cable sheathing 0.10 inch in thickness in about 8 years.[3] Tin, although resistant to corrosion in air and pure water, is severely corroded by alkalies, and aluminum is attacked by both alkalies and acids. The

noble metals such as gold, silver and platinum, being less reactive chemically than the more basic metals, are as a group the least corrodible, yet silver tarnishes markedly in moist atmospheres containing volatile sulfur compounds; gold is attacked by halogens in solution and platinum by fused alkalies.

The protection of metals from corrosion may be accomplished in general either by maintaining a non-corrosive surrounding environment or by coating the metallic parts with paints, lacquers or more corrosion-resistant metals. Such measures as the control of humidity and dust in interior atmospheres, deoxygenation of boiler waters, the use of passivators such as chromates, carbonates, phosphates, silicates and alkalies in the water-cooling systems or water scrubbers of air conditioning equipment and the use of cathodic protection which consists in setting up an electrolytic cell in which the metallic part subject to corrosion is made the cathode, are typical examples of environmental control designed to inhibit corrosion. Another well-known example of avoiding corrosion by control of environment is the protection of underground cables which is afforded by the proper drainage of electrical stray currents which have been picked up by the cable network. Where it is infeasible to maintain an inert environment the use of non-ferrous metallic coatings is of great value in the preservation of steel products. Much metallurgical work has been devoted in recent years to the development of corrosion-resistant alloys. In both of these cases the protective feature consists in a naturally developed surface film. Where natural films afford insufficient protection it becomes necessary to resort to coatings of organic materials such as paints, lacquers, enamels, complex structures of such materials as pitches or asphalts with jute felt or paper, etc. It has been estimated that one hundred and twenty million gallons of paint are used annually for corrosion prevention.[4]

Corrosion may be defined in most general terms as the chemical reaction of a metal with the non-metallic constituents of its environment. In this sense any reaction in which a metal is degraded to one of its compounds, such as an oxide, hydroxide, acid or salt is a corrosion reaction. The nature of the reaction which occurs in any given case depends both upon the reactivity of the metal, its purity, physical state and surface condition and upon the character of the environment, that is, upon the chemical components present, their physical phases and concentrations. It also depends upon the temperature. Corrodibility is not wholly an inherent property of a metal which can be determined by a single arbitrarily chosen corrosion test of any sort; even the relative order of corrodibility of a series of metals

is not constant.[5] For example, iron, magnesium and zinc corrode in conductivity water exposed to the air in the order given; in sodium chloride solution the order is magnesium, zinc and iron, while finally in strong alkali solutions these metals corrode in the order: zinc, iron and magnesium.

It is apparent that the occurrence of corrosion depends upon both the character of the environment and of the metal. While the environment in which a metal is used is usually complex, it is generally possible to recognize those constituents which exert a controlling influence on the course of corrosion. In the ordinary atmosphere water vapor and oxygen are major factors in the process. Other substances such as sulfur dioxide, chloride ions and dust also influence the character of the corrosive attack. The green patina, a basic sulfate, which forms in the course of time on copper exposed to the air and which may preserve the metal for centuries, owes its origin to traces of sulfur dioxide in the atmosphere.[6] On the other hand, the higher concentrations of sulfur dioxide prevailing in the neighborhood of smelters which treat sulfur bearing ores may rapidly corrode copper telephone wires to destruction. Steel containing up to 0.25 per cent copper is about two-fold more resistant to corrosion than non-copper bearing steel in most industrial atmospheres; in New York City, however, the very small chloride ion content of the atmosphere, an otherwise typical industrial atmosphere, largely destroys the corrosion resistance conferred by the copper. Rainfall is an influential factor in the corrosion of metals exposed to the atmosphere. It may increase corrosion by removing soluble corrosion products from the surface of the metal, or it may retard corrosion by washing away dust particles and electrolytes, both of which promote corrosive attack. For example, in New York City the daily application of a water spray increased the rate of corrosion of zinc by 30 per cent but decreased the rate of corrosion of different ferrous materials from 30 per cent to 46 per cent, the amount of corrosion being determined by loss of weight measurements. The corrosion products of zinc are appreciably more soluble than those of iron and presumably were largely removed by the frequent washing. On the other hand, the deliquescent nature of the corrosion products of iron [7] at humidities prevailing a large part of the year provide, in contrast to the relatively less deliquescent corrosion products of zinc, a film of moisture more or less saturated with corrosive electrolytes and dust particles. This film is diluted or otherwise removed by the water spray.

Indoor environments differ from the outside atmosphere mainly in being drier, cleaner and subject to less pronounced temperature varia-

tions. As a consequence, metals corrode somewhat less rapidly indoors but the character of the attack generally resembles that shown out of doors. Underground exposures to soils and waters are often severe and cause ferrous structures to fail unless protective non-metallic coatings are employed. Considerable progress has been made within the past ten years in determining the corrosivity of soils and developing adequate preservative coatings.[8]

The best measure of the tendency of a metal to corrode is, in thermodynamical terms, the decrease in free energy which accompanies the chemical reaction involved in the process, i.e., the difference in energy between the initial and final state of the system. This may be obtained by simple calculation and is of value in showing whether or not it is possible for corrosion to occur under the conditions defined. There is no assurance however, that reactions which are possible will actually take place within a reasonable time, if at all. Calculations [9] show that, exposed to the atmosphere containing moisture, aluminum, zinc, tin, iron, nickel, copper and silver may corrode to their respective hydroxides. If, oxygen be excluded the last three metals listed cannot corrode and iron only to the lower state of oxidation. These calculations, however, give no information as to the rate of corrosion or the mechanism by which it takes place, matters of great practical importance.

The rate of corrosion or of any other chemical reaction bears no direct relationship to the energy changes involved; it cannot be predicted but must be measured in some form.[10] Obviously the rate of corrosion depends upon the nature of the chemical reactions at the surface of the metal. Generally, secondary reactions are involved and the slowest step in the process controls the rate. The limiting factor is usually some sort of barrier,—a film of gaseous or solid corrosion products at the surface.

The mechanism by which corrosion occurs may be one either of direct combination of the metal and non-metal or the replacement by the corroding metal of hydrogen or another metal in compounds. The oxidation of metals, particularly at higher temperatures, halogenation reactions, such as the chlorination of aluminum, and the reaction of copper and sulfur, are examples of direct combination. In many of the reactions which occur in the atmosphere, such as the formation of tarnish films, the processes are somewhat obscure. When zinc corrodes in the ordinary atmosphere an oxide film forms in the early stages which is pseudomorphic with the metal [11] but which is converted eventually into the ordinary granular form of zinc oxide. The rate of corrosion of zinc is determined by the rate of this conversion

process,[12] the granular oxide film having no retarding influence. This is shown by the linear relationship in Fig. 1 which compares the corrosion-time curves for zinc, copper, lead, and iron in the atmosphere. In this case corrosion is expressed in terms of weight gain due to the accumulation of corrosion products. It will be observed that the relationship for copper is a parabolic one, indicating that the process is controlled by the rate of diffusion of oxygen through the increasingly thicker oxide film. Up to a thickness of about 10 Å the film is invisible and when formed in pure air is impervious to volatile sulfur compounds.

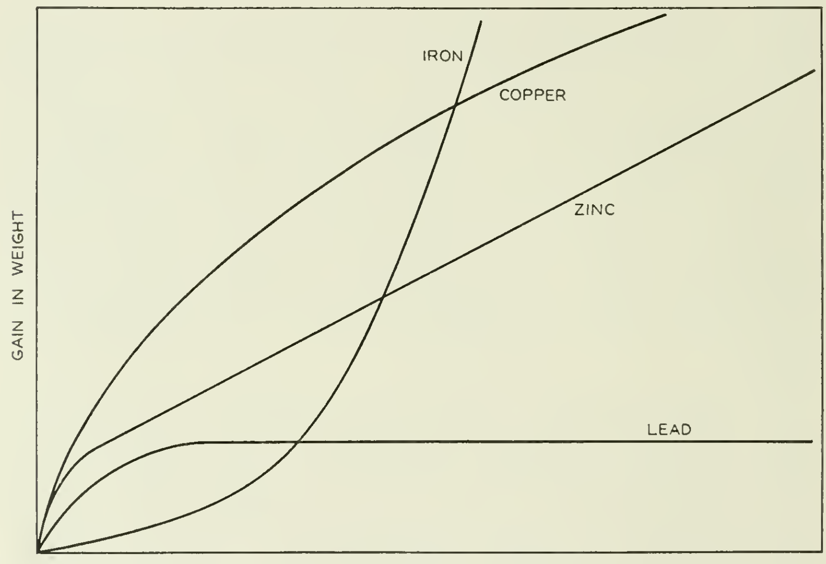


Fig. 1—Corrosion-time relationships characteristic of certain metals exposed to the atmosphere.

A third type of corrosion-time curve represented by lead becomes parallel with the time axis after the initial stages. Evidently the film in this case is impervious to the constituents of the environment. The curve given for iron indicates that the film which forms in the initial stages of the exposure exerts an accelerating influence upon the subsequent rate of oxidation.

It should be mentioned that this acceleration occurs only at humidities above what has been called the "critical" humidity. By this term is meant the relative humidity corresponding to the vapor pressure of a saturated solution of the corrosion products, which depends upon the composition and to some extent the structure of these

products. For iron this is somewhere between 40 and 65 per cent relative humidity, probably nearer the latter figure.[13] The atmospheric corrosion products of the non-ferrous metals are in general less deliquescent, i.e., have higher critical humidities; for nickel it appears to be above 70 per cent relative humidity and for zinc and copper above 75 per cent. It has been suggested that the marked increase in the resistance of copper containing about 0.5 per cent arsenic to atmospheric corrosion is due to the fact that the presence of arsenic renders the corrosion product less hygroscopic.[14] The effect of copper in copper bearing steel may be of the same nature;[15] apparently the inhibiting action in this case does not lie in the production of a film which is any more resistant to attack initially than that on ordinary steel.[16] In dust-free air, even at high humidities, iron does not corrode but forms an invisible protective film. Electron beam studies[17] have indicated the structure of this film to be a form of ferric oxide which has been designated as α Fe_2O_3 in contrast to the composition of a non-protective form which appears to be γ $FeO \cdot OH$.

The presence of dust particles in the atmosphere greatly increases the rate of corrosion of iron. In this case, as well as in the accelerated attack which occurs above the point of critical humidity, the process involves the displacement of hydrogen from water. Other common examples of this type are the solution of metals in acids or alkalies, the reaction of sodium with water, the deposition of metallic copper from copper sulfate solution by metallic zinc and in general the corrosion of metals in moist air, in soils and in water. It is well established that the process of corrosion in these cases is electrolytic in character, i.e., that corrosion occurs by means of the operation of small galvanic cells at the surface of the metal. The primary reactions of these cells may be and generally are followed by important secondary chemical reactions of the products of electrolysis with the constituents of the environment. Between the anode and cathode areas there is a flow of current through intervening electrolytic paths of greater or lesser resistance. Naturally the amount of corrosion is proportional to the amount of current flow.[18] It is largely the distribution and size of anode areas which determines the character of the corrosive attack. For a given amount of metal dissolution the existence of relatively few anode areas small in size obviously leads to pitting, while if there are numerous anodes uniformly distributed, corrosion will likewise be uniform. The distribution of anodes is determined by the inhomogeneity of the base metal, the character of the films which are formed, the accidental contact of inert bodies and the conductivity of the surrounding electrolytic media.

The potential differences which are responsible for the existence of corrosion cells arise either from some chemical or physical inhomogeneity of the metal or from some inhomogeneity of the environment at the metal surface. These cells or galvanic couples provide the means whereby the natural tendency of metals to corrode may express itself. The nature of the electrodes composing some of these electrolytic cells is given in the following table:

Cathode	*Anode*	*Example of Cell*
Noble metal	Base metal	Two-phase alloys, metal containing metallic impurities, two metals in contact or metal with porous metal coating.
Metal oxide	Metal	Iron with porous oxide scale.
Dust	Metal	Iron with carbonaceous dust particles on surface.
Metal freely exposed to oxygen	Metal in diminished oxygen supply	Exclusion of air at point of contact of metal and inert material or decreased oxygen concentration at bottom of pits on surface of metal.
Metal in concentrated acid, alkali or salt solution	Metal in dilute acid, alkali or salt solution	Metal in contact with solution of two different concentrations or two different solutions.
Annealed or coarsely crystalline metal	Same metal strained or of small crystal size	Metal which has been subject to non-uniform heat treatment or cold working.

It will be observed that there are ample possibilities for a metal of even high purity to corrode. The existence of more than one metallic phase in metals in industrial use often does not have a significant bearing upon their corrosion behavior. For example, studies conducted by these Laboratories have shown that high purity lead, lead hardened with 1 per cent antimony, 3 per cent tin or 0.03 per cent calcium, when used as cable sheathing show approximately the same resistance to corrosion. While one of these materials may be somewhat more corrodible than others in a given natural environment, the reverse will be true for another set of conditions. The environment is of far greater importance than the composition of the sheathing and consequently the control or avoidance of corrosion is attained by maintaining the cable plant in non-corrosive surroundings.

Of the types of corrosion cells listed above, those due to concentration differences in the surrounding medium are among the most prevalent. Differential aeration is one of the most common causes of corrosion. Lead is corroded beneath the point of contact with a large grain of sand when in an atmosphere containing oxygen and water vapor.[19] The point of contact is less accessible to oxygen than surrounding parts. The potential of this cell is somewhat less than 0.1 volt, the value determined in these Laboratories for the difference between the

potentials of lead in identical solutions exposed respectively to oxygen and to an inert gas, nitrogen. An even more striking example of the operation of oxygen concentration cells is observed when glass beads are in contact with a lead surface wet by a film of sodium chloride solution exposed to the air.[20] After a period of a few months a ring of bright red corrosion product forms around an anodic pit located directly beneath each bead. This red corrosion product, a red tetragonal form of litharge, is characteristic of an alkaline attack. Apparently sufficient caustic soda to cause corrosion (a solution of pH 12 approximately) was produced by the differential aeration cell resulting from the contact of the bead with lead.

The foregoing paragraphs have described some of the complexities encountered in corrosion processes. In view of these complexities it has been one object of these Laboratories for some years to advance the development of a generalized theory of corrosion applicable to all cases of corrosion of the replacement type, since it is this type of process which prevails in atmospheric, soil and water exposures. Direct combination of metals with non-metallic elements is limited largely to somewhat extreme conditions such as those of industrial processing. Attention will now be directed to the general theory of corrosion in its present state of development.

The fundamental reactions of corrosion processes of the replacement type as represented by the operation of corrosion cells are as follows:

$$\text{At the anode: } M \xrightarrow{-e} M^+;\ 2(\mathrm{OH})' \xrightarrow{-2e} \mathrm{O} + \mathrm{H_2O}.$$

That is, the metal sends ions into the solution or there are plated out non-metallic elements such as oxygen. Either process is accompanied by a loss of electrons. If corrosion is induced by an externally applied potential, as for example when stray electrical currents flow from underground metallic structures to earth, oxygen atoms may leave the surface in the form of molecular oxygen.

$$\text{At the cathode: } M^+ \xrightarrow{e} M;\ \mathrm{H}^+ \xrightarrow{e} \mathrm{H};\ 2\mathrm{H} \rightarrow \mathrm{H_2}\text{ gas}.$$

This process consists in plating out either metal or hydrogen atoms. In the latter case atomic hydrogen either leaves the metal surface as hydrogen molecules or acts as a reducing agent, being in turn oxidized. The electrode reactions may be combined in equations as follows:

1. Solution of the metal: $M + \mathrm{H}^+ \rightarrow M^+ + \mathrm{H}$.
2. Removal of hydrogen: (*a*) $2\mathrm{H} \rightarrow \mathrm{H_2}$,
   (*b*) $2\mathrm{H} + \mathrm{O} \rightarrow \mathrm{H_2O}$.

The corrosion reaction represented by the first reaction automatically stops when the metal surface becomes covered with hydrogen atoms and can only proceed when and as hydrogen is removed by one of the processes above given. Some metals evolve molecular hydrogen if the potential of the corrosion cell is slightly over that required to plate out hydrogen atoms. In the greater number of cases, however, molecular hydrogen is not liberated rapidly, i.e., as bubbles, unless the potential is substantially higher than that required to plate out the atomic form. The additional potential required to evolve molecular hydrogen against the normal atmospheric pressure is termed "overvoltage." In the case of metals of high hydrogen overvoltage, oxygen or oxidizing substances are required to facilitate the removal of the hydrogen film if corrosion is to proceed at an appreciable rate. Consequently corrosion may be controlled by the rate at which oxygen reaches the metal surface and "depolarizes" it. In many cases both processes are operative. For example, when iron is corroded in dilute potassium chloride solutions it is interesting to note that under a pressure of one atmosphere of oxygen 1/13 of the total corrosion is accompanied by the discharge of hydrogen gas and 12/13 by the oxidation of hydrogen to form water. When the oxygen pressure on the system is raised to 25 atmospheres by conducting the experiment in a closed bomb the total rate of corrosion is increased 45-fold owing to the increased rate of hydrogen oxidation, the rate of hydrogen evolution being practically unaffected. In practical cases, as would be expected, either the character of the corroding medium or the purity of the metal may affect this ratio of hydrogen control to oxygen control. In tests in which iron specimens were totally immersed in sea water exposed to oxygen, about 35 per cent of the total corrosion was accompanied by hydrogen evolution, whereas for similar specimens in a half-normal solution of pure sodium chloride (which corresponds roughly to the salt content of sea water) only 5.6 per cent of the corrosion was of the hydrogen evolution type.[21] The presence of one part per million of impurity in zinc considerably accelerates the rate of corrosion [22] mainly by stimulation of hydrogen evolution.

The intensity with which a metal tends to send metal ions into solution increases with the basic character of the metal. The presence of ions of the metal in the solution may be considered to constitute a force opposing this ionization tendency and the value of the resulting equilibrium is known as the potential of the metal in that solution. The value is constant or static only so long as there is no flow of current between the metal and solution. The molal potentials and normal potential of metals are their static potential in solutions of

their salts in which the metal ion concentrations are molal and normal respectively. A table of the values of such potentials constitutes the so-called E.M.F. series. This comparison of metals is of little value in predicting either the driving force or the rate of operation of practical corrosion cells, the electrodes and ionic concentrations of which bear little correspondence to those defined in the E.M.F. series. Moreover, the practical case is complicated in many instances by the gas electrode behavior of the metal and by the flow of current during the corrosion process. For example, the potentials of many metals when measured in the atmosphere are much more noble or cathodic than might be expected from a knowledge of the E.M.F. series. This is due to the fact that in the presence of moisture and oxygen, metals may function wholly or in part as oxygen electrodes. The exact values of these electrodes depend upon the concentrations of oxygen present, and upon the acidity of the solution. This explains the origin of the potential difference of the differential aeration cells to which reference has been made previously.

If the corrodibility of copper in the presence of moisture were judged solely from the position of the metal in the E.M.F. series no attack would be expected, since in this series copper is more noble than hydrogen, the element which must be displaced in the corrosion process. As a matter of fact copper does not corrode even in hydrochloric or sulfuric acids in the *absence of available oxygen.* It is readily corrodible, however, in nitric acid because in effect under these circumstances the position of the hydrogen electrode is rendered cathodic to copper (i.e., more noble than copper) owing to the depolarizing influence of oxygen. It is probably for this same reason that oxygen markedly accelerates the corrosion of monel metal in 3 per cent sulfuric acid.[23]

The change in electrode potential with current flow, polarization, may be illustrated in a simple experiment as follows: If the zinc coating is removed from a portion of the surface of a strip of galvanized iron, exposing thereby the underlying iron surface, one has what amounts to a simple galvanic zinc-iron couple. If this couple is completely immersed in a dilute salt solution and potential measurements are made at different points on the iron and zinc surfaces by means of a calomel half-cell, it will be observed that the potentials of iron and zinc at some distance from the iron-zinc interface are approximately those values obtained for these metals separately in the same electrolyte; while, on the other hand, the value of the iron potential near the interface has become more anodic and the potential of zinc near the interface has moved in the cathodic direction, i.e., the difference in potential between iron and zinc near the interface of the two metals is appreci-

ably less than when taken at points more widely separated. Figure 2 gives a schematic representation of this experiment. The change in potential is the result of current flow through the electrolyte from zinc to iron. The current densities are highest in the region of the interface, the metal ion concentration becoming increased at the anode area and decreased at the cathode area, producing thereby anodic and cathodic polarization, respectively.

This polarization behavior of corrosion cells largely determines the rate of corrosion. It is obvious that the effective potentials of corrosion cells may be reduced by polarization to zero, in which case the rate of corrosion is limited to that required to maintain this polarization.

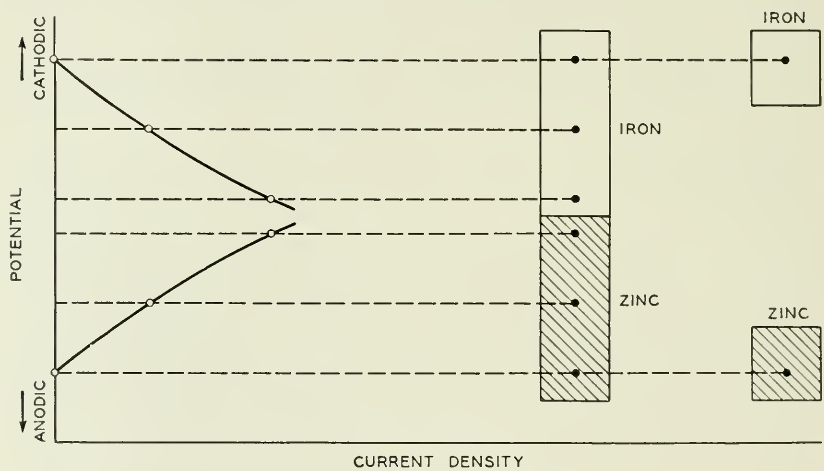


Fig. 2—Illustration of galvanic polarization.

tion. In other words, the progress of corrosion may be controlled by the extent either of anode polarization or cathode polarization or both, that is, either one may determine the final result. Figure 3 represents the variety of current density-potential relationships which may exist in corrosion cells. In Cell 1, in which there is no appreciable polarization of either anodic or cathodic areas (as indicated by the small change of potential with current), corrosion current flow is limited by the resistance of the electrolytic paths between anodes and cathodes and since this may be small if these areas are contiguous the corrosion rate may be high. In Cell 2 the anode is highly polarized as represented by the solid line or progressively less polarized as the point of intersection with the non-polarized cathode occurs at higher and higher current densities as represented by the dotted lines. In a

similar way Cell 3 shows cathode polarization only and Cell 4 both anode and cathode polarization. Since the rate of corrosion is proportional to the flow of current per unit area it is obviously limited in the last three cases by the values of current density at which the polarization curves intersect. In the presence of adequate oxygen or in cases where hydrogen is readily discharged corrosion cells are likely to resemble Cell 1. Where this is not the case, as in the absence of oxygen or where the cathodes have high values for hydrogen overvoltage, the result will be as shown for the cathodically polarized Cell 3. The presence of an inhibitor such as a positively charged colloid or

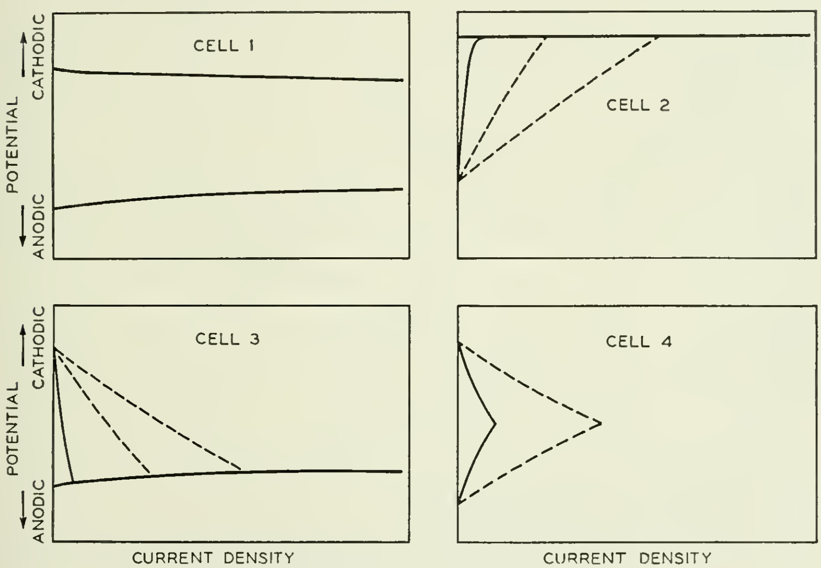


Fig. 3—Types of polarization in corrosion cells.

the amalgamation of the metal surface with mercury are other conditions which promote cathode polarization. On the other hand, the action of passivating agents such as chromates, silicates and in some cases sulfates, carbonates, etc., is to produce anodic polarization as in Cell 2. It will be observed that whereas in the presence of inhibitors of the type mentioned above which polarize the cathode the resulting potential of the corrosion cells and therefore of the metal specimen as a whole should move in the anodic direction as the process of inhibition takes place, in the case of passivating agents (which influence anode processes) the effect of increasing passivation is a trend of potential in the cathodic or noble direction. In both cases corrosion is retarded or prevented entirely.

The manner in which the conductance of the surrounding electrolyte influences the rate of corrosion is illustrated in Fig. 4A in which the upper curve represents the cathodic and the lower the anodic polarization. Assuming equal anodic and cathodic areas the corroding current density for the lower conducting solution is represented by "M" and that for the higher conducting solution by "N." In the actual case where anodes and cathodes are in close juxtaposition, the internal resistance is low and consequently the corroding current density approaches that represented by the intersection of the polarization curves.

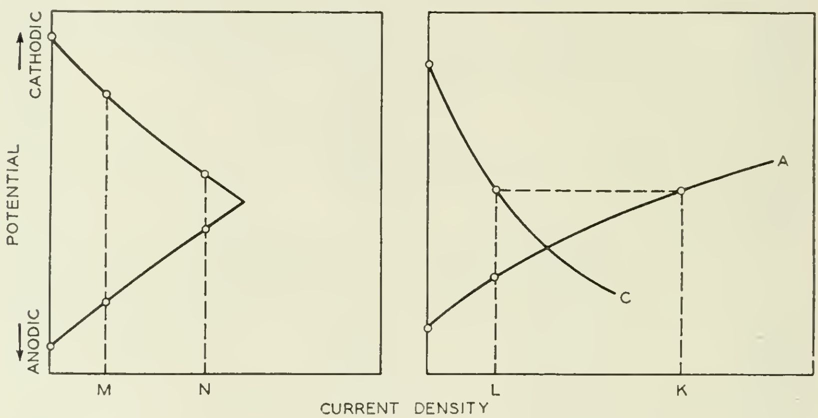


Fig. 4—Effect of conductance and of electrode area on corrosion current densities.
M = Lower conducting solutions.
N = Higher conducting solutions.
L = Corrosion current density for cells of equal cathode and anode areas.
K = Corrosion current density when ratio anode area to cathode is small.

Thus far consideration has been confined for the sake of simplicity to corrosion cells in which the anodic and cathodic areas are equal. Usually in actual experience this is not the case. In corrosion characterized by pitting, the anodic area is generally small compared to the cathodic areas. This situation is illustrated in Fig. 4B in which it will be seen that under these conditions a high corroding current density corresponding to a rapid rate of attack may occur. Conversely, in cases where the ratio of anode areas to cathode area is large, the rate of attack will be small, being thus controlled by cathodic polarization. In this connection it may be of interest to consider the effect of impurities upon rate of corrosion. If the contaminating metal is anodic and exists as a separate phase it will tend to dissolve with the formation of small pits which having once formed may possibly

continue to function as the anodes of oxygen concentration cells. If, on the other hand, the metallic impurity is cathodic and present as a separate phase, corrosion will be rather more uniform in character and its rate will be controlled in the absence of oxygen by the ability of the impurity to discharge hydrogen. Unless its overvoltage is low, that is, unless it discharges hydrogen readily, the rate of corrosion will be slow, the corrosion cells being polarized cathodically. The presence of oxygen or oxidizing agents under these conditions will depolarize these cathodic areas and accelerate corrosion.

From the foregoing it is apparent that a knowledge of the anodic and cathodic current density potential relationships which are established on the surface of a metal in a given environment would make possible an understanding of the processes which are taking place and lead to a prediction of corrosion behavior. It is generally impossible to measure these quantities as they relate to individual corrosion cells owing to a lack of knowledge of the electrode areas involved. Probably these comprise a wide range of sizes and change in size with the progress of corrosion. Sometimes the nature of the cathodes is also uncertain. Practically, however, it is a simple matter to determine a composite of the resultant potentials and their change with time. These time-potential measurements indicate whether the process is anodically or cathodically controlled and in some cases furnish information as to the rate at which it is proceeding, experimental facts which are of value in predicting corrodibility. A recording potentiometer is of considerable assistance in this connection.

Figure 5 illustrates schematically the correlation between these time-potential relationships and the anodic and cathodic polarizations which determine their positions. It will be seen that the potential of iron in a solution of potassium sulfate (represented by the lower solid curve) is mainly determined by the anodic potential of iron in the solution, the cathodic areas being polarized. When potassium chromate is added to the solution the resultant potential of iron is shifted markedly in the cathodic direction, the position being determined by anodic polarization. The actual values of the potential of iron in these cases are of the same order as that of iron alloyed and rendered passive by the addition of chromium and nickel.[24]

In a solution containing hydrogen peroxide, iron is passive even in the acid range as shown in Fig. 6. The abrupt cathodic shift in the potential of iron in the region of pH 6.5–6.8 also shown in Fig. 6 is due evidently to a film which affects both anodic and cathodic areas and which judged from this position would be expected to be less protective than the more pronouncedly passive films (indicated by

their cathodic positions) produced by hydrogen peroxide and chromate ions. A trend of potential in the anodic direction with time, while it may suggest the breakdown of a passive film, does not necessarily indicate the onset of a corrosive attack, for should there also be present substances such as positively charged colloids or other products [25] which tend to raise and maintain hydrogen polarization at the cathodic areas, the metal may suffer little or no attack.

Whether the films which form in corrosion processes are protective in character depends to a considerable extent upon their position or location with reference to the surface of the metal and this in turn

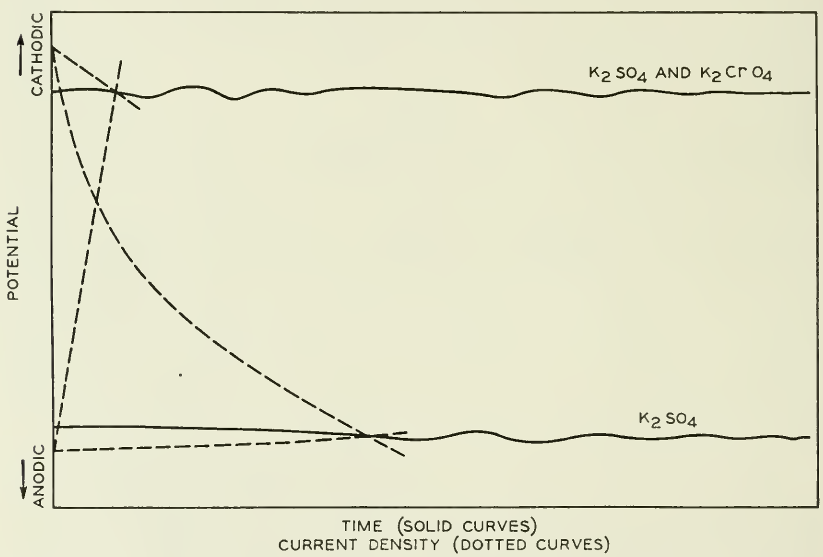


Fig. 5—Time-potential relationship of iron in K_2SO_4 as affected by addition of K_2CrO_4.

depends upon the solubility of the corrosion products in the medium adjacent to the surface. When iron corrodes in the presence of moisture, ferrous ions are produced at anodic areas and hydroxyl ions at cathodic areas, the process continuing until the solubility limit of ferrous hydroxide is attained, whereupon this compound begins to be precipitated as a gelatinous film over the surface of the metal. Increasing the alkalinity of the environment naturally represses the solubility of this compound, precipitating it with less solution of iron. In the absence of oxygen the ferrous hydroxide film tends to inhibit corrosion by maintaining hydrogen polarization. When, on the other hand, oxygen is accessible to the system, ferrous ions are oxidized with the result that ferric hydroxide being less soluble than ferrous

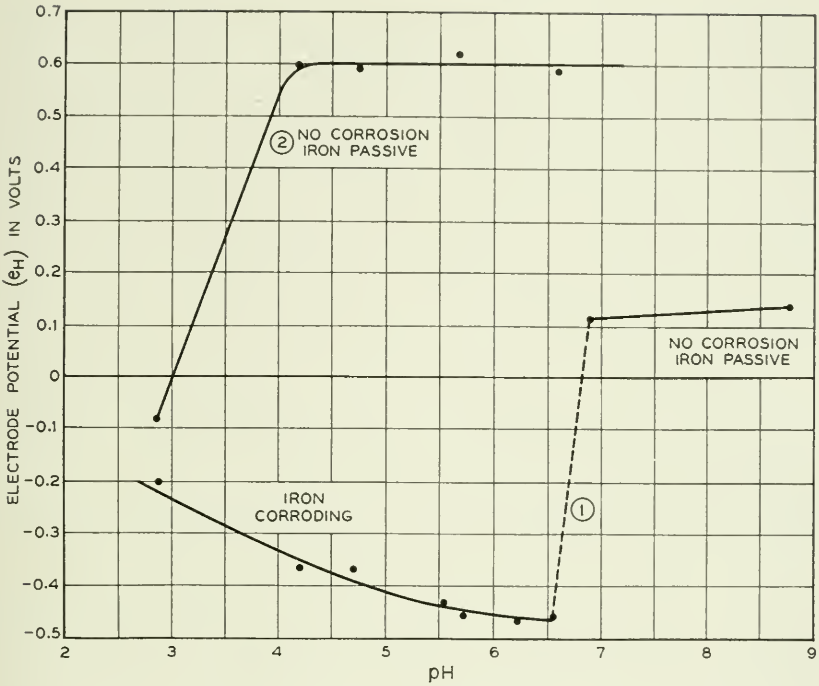


Fig. 6—Effect of acidity and of hydrogen peroxide on the equilibrium potential of iron in buffer solutions of varying acidity.

1. Potential of iron in tenth normal solutions of sodium acetate-acetic acid solutions.
2. Potential of iron in tenth normal solutions of sodium acetate-acetic acid solutions + 0.6 per cent hydrogen peroxide.

hydroxide is precipitated and forms a more or less porous rust film at an appreciable distance from the metal surface. Owing to the mildly amphoteric nature of iron there may exist, especially under alkaline conditions or higher temperatures a considerable concentration of ferrite (FeO_2') ions which, upon reacting with ferrous ions, may precipitate ferrous ferrite (Fe_3O_4) or black magnetic oxide of iron which, also being precipitated in a somewhat granular form at some distance from the surface of the metal, is non-protective. In contrast to these examples is the highly protective film of silicate which presumably forms upon lead and lead rich alloys when immersed in water or soil solutions containing as little as ten parts per million of silicate. As is well known, distilled water is corrosive to these metallic materials. Were it not for this fortunate effect of silicates upon lead it is doubtful that it or its alloys could be used for cable sheathing in the present type of underground construction which permits exposure to soil and surface waters at times.

Studies have shown that the points of failure of air-formed films on iron and steel surfaces indicated by the initial appearance of anodic or rust spots depends upon the previous history of the specimen and upon the medium in which the test is conducted.[26] For example, an increase in time of pre-exposure to oxygen or exposure to higher temperatures decreases the number of initial anodes, while increasing the chloride content of the medium or the presence of sulfide on the metal surface [27] increases them. Whether corrosion continues at the points of initial attack often depends upon the self-healing ability of the film, that is, upon plugging the fissures or pores in the film with corrosion products.

Various methods have been considered for the determination of the quality of protective films. In the case of aluminum and its alloys the amount of leakage current which may pass through anodically formed films throws some light upon resistance to corrosion. Another promising method applied to iron steel and alloy steels has been to determine by potential measurement the amount of chloride required to destroy passive films formed in water or chromate solutions.[28]

The rate of film formation is in a sense a measure of the activity of a metal surface, that is, a measure of the rate at which a metal might corrode in a homogeneous environment in the absence of film forming constituents. For example, it will be seen in Fig. 7 that the potential

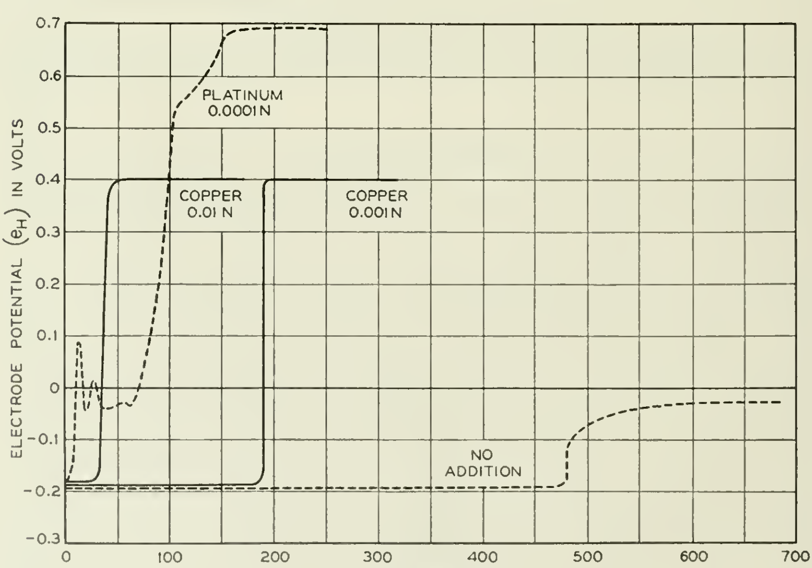


Fig. 7—Effect of traces of copper and platinum on the potential of lead in 0.1 N H_2SO_4.

of lead in tenth normal sulfuric acid breaks rather abruptly after about 500 minutes to the potential of a gas electrode. Apparently the anodic areas become progressively covered with a film or sulfate until substantially the entire surface is passive. Upon the introduction of a drop of thousandth normal copper sulfate, the sulfation of a similar lead specimen is consumated in about 200 minutes at the end of which time the potential breaks to that of the cathodic areas of copper which have been formed by replacement deposition. A still higher concentration of copper brings about sulfation still more rapidly and when the solution is contaminated with platinum a break to the potential of platinum occurs after a still shorter period. In the same manner, the relative corrodibilities of leads of various purities and certain lead alloys [29] has been compared.

The foregoing discussion of the application of electrochemical methods to corrosion investigations outlines techniques by means of which it is possible to get information of the following kind. By the position of the potential of a metal against its environment and the trend of this potential with time it is possible to determine whether the corrosion process is controlled by reactions occurring at the anodic areas, the cathodic areas or both, that is, whether there is a tendency toward passivity, inhibition or progressive attack. Measurements of film stability whether in terms of the leakage current which may be passed through the film, or in terms of the amount of film-forming material (such as chromates) required to produce passivity or the amount of film destroying material (chlorides) required to render the metal active, furnish information as to the quality of corrosion resistant films. Finally measurements of the rate at which a film forms on a metal when placed in a film-forming environment also throws light upon the relative surface reactivity of the metal. Such information is of assistance in determining the rates of corrosion in homogeneous corrosive environments or the rate of passivation in film-forming environments. It is evident in all of these cases that the interpretation of the experimental data which are obtained and the application of the findings to practical corrosion problems is considerably facilitated by a chemical knowledge of the environments in which metals are used as well as the composition and physical state or structure of the metallic material. With such measurements and such knowledge it is possible to predict corrosion behavior and to obtain an understanding of corrosion problems usually not possible by ordinary empirical corrosion tests.

To summarize, the process of corrosion may be one of direct combination of a metal and a non-metal or it may be one in which hydrogen

or another metal is displaced from the medium at the surface of the metal. In either case reaction products appear and usually exert a controlling influence upon the progress of attack. In the replacement type of corrosion process, in which the attack occurs by means of the operation of small galvanic couples at the surface of the metal, it is possible to consider separately those influences which affect anode behavior and those involved in cathode behavior. The course of corrosion or resistance to corrosion may be explained in terms of the anode or cathode control of the process. It is apparent then that a knowledge of both the composition and condition of a metal surface and of the surrounding environment is requisite to an understanding of corrosion problems.

References

1. Speller, *Tech. Pub. No.* 553C, *A. I. M. E.*, February 1934.
2. Hudson, *Trans. Faraday Soc.*, *25*, 177 (1929).
3. Logan and Taylor, *Bur. Stds. Jour. Research*, *12*, 119 (1934).
4. Speller, loc. cit.
5. Bengough, *Jour. Soc. Chem. Ind.*, *52*, 195 (1933).
6. Moore and Liddiard, *Jour. Soc. Chem. Ind.*, *54*, 786T (1935).
7. Patterson and Hebbs, *Trans. Faraday Soc.*, *27*, 277 (1931).
8. Logan, *Trans. Electrochem. Soc.*, *64*, 113 (1933); Ewing, *Am. Gas. Assn.* Report on Pipe Coatings and Corrosion (1933); Dennison, *Bur. Stds. Jour. Research*, *7*, 631 (1931); *Am. Gas Jour.*, Sept. 29, 1932; Dennison and Hobbs, *Bur. Stds. Jour. Research*, *13*, 125 (1934).
9. Forrest, Roetheli and Brown, *Ind. & Eng. Chem.*, *23*, 650 (1931).
10. Johnston, *Ind. & Eng. Chem.*, *26*, 1238 (1934).
11. Finch and Quarrell, *Proc. Phys. Soc.*, *46*, 148 (1934).
12. Finch, *Discussion Trans. Faraday Soc.*, *31*, 1116 (1935).
13. Vernon, *Trans. Electrochem. Soc.*, *64*, 31 (1933).
14. Vernon, *Trans. Faraday Soc.*, *27*, 225 (1931).
15. Borgmann, *Bell Lab. Record*, *10*, 233 (1932).
16. Mears, Sec. D Part 2, *Third Report of Corrosion Comm. of Iron and Steel Inst. and British Iron and Steel Federation* (1935).
17. Rupp, *Kolloid-Z.*, *69*, 369 (1934).
18. Evans, Bannister and Britton, *Proc. Roy. Soc.*, *A131*, 355 (1931).
19. Burns and Salley, *Ind. and Eng. Chem.*, *22*, 293 (1930).
20. Experiments carried out by Mr. H. S. Phelps of the Philadelphia Electric Company, through whose courtesy the author was permitted to inspect the results.
21. Bengough, *Jour. Soc. Chem. Ind.*, *52*, 229T (1933).
22. Harris, *Trans. Am. Electrochem. Soc.*, *57*, 313 (1930).
23. Searle and Worthington, *Int. Nickel Co. Bull.* T-3.
24. Bannister and Evans. *Jour. Chem. Soc.* 1361, June 1930.
25. Herzog, *Trans. Electrochem. Soc.*, *64*, 87 (1933).
26. Mears and Evans, *Trans. Faraday Soc.*, *31*, 527 (1935).
27. Mears, loc. cit.
28. Fenwick, *Indus. and Engg. Chem.*, *27*, 1095 (1935).
29. Haring and Thomas, *Trans. Electrochem. Soc. Preprint* 68–13 (1935).

Magnetic Measurements at Low Flux Densities Using the Alternating Current Bridge

By VICTOR E. LEGG

A resumé is given of the basic relations between the magnetic characteristics of the core of a coil and the inductance and resistance of the coil as measured on an alternating current bridge. Modifications of the simple relations to take account of the interactions of eddy currents and hysteresis in the core material are developed, and are seen to require a more complicated interpretation of the data in order to obtain an accurate separation of the eddy current, hysteresis, and "residual" losses. Means are described of minimizing or eliminating the disturbing effects of distributed capacitance, leakance and eddy current loss in the coil windings. Essential details of the alternating current bridge and associated apparatus, and of the core structure, are given.

THE modern alternating current bridge, with its high precision and sensitive balance, has almost completely superseded the ballistic galvanometer for determining the magnetic properties of core materials at the low flux densities employed in telephone and radio apparatus. The suitability of the alternating current bridge for this purpose has been recognized for some time,[1] but the continued improvements in magnetic materials, and the more exacting requirements of modern communication apparatus, have necessitated refinements in apparatus, in technique, and in interpretation of measurements. This paper considers the modified technique required to take account of eddy current shielding and hysteresis in the magnetic core, distributed capacitance and leakance in the coil winding, and the necessary details of the bridge and associated apparatus to realize the desired accuracy of measurements.

Fundamentally, the a-c. method involves measurements of the inductance and effective resistance of a winding on the test specimen, such measurements being made at several frequencies, and at several values of current.[2] From these measurements the magnetic properties of the test core can be computed for the low flux density range. The details of such calculations will be given below, beginning with approximate methods, and proceeding to successively more accurate computations.

[1] M. Wien, *Ann. d. Physik* [3] **66**, 859 (1898).

[2] The annular form of magnetic core, wound with a uniformly distributed test winding will here be treated, but the results will be found to be readily transferable to other forms of core.

Simple Analysis of Inductance; Hysteresis Neglected

The magnetizing force in a thin annular core of mean diameter d (cm.) due to current i (ampere) flowing in a uniform winding of N turns is

$$H = \frac{0.4Ni}{d} \text{ œrsted.} \tag{1}$$

In a core of appreciable radial thickness, the effective magnetic diameter rather than the arithmetical mean diameter must be used in this and following equations, as will be explained in eq. (61).

When the bridge is balanced with sinusoidal current of peak value i_m, the peak inductive voltage drop across the standard coil, $2\pi fLi_m$, must equal that across the test coil $2\pi fN\Phi_m \times 10^{-8}$, where Φ_m is the peak magnetic flux in the coil.[3] Whence, for an annular coil,

$$L = \frac{4N^2\Phi_m}{H_m d} \times 10^{-9} \text{ henry.} \tag{2}$$

The flux within an annular coil is composed of that in the core and that in the air space. The expression for inductance can therefore be separated into two terms, giving

$$L = \frac{4N^2}{d}(\mu_m A + A_a) \times 10^{-9} \text{ henry,} \tag{3}$$

where A and A_a are the cross-sectional areas of the core and residual air space, respectively, and μ_m is the magnetic permeability of the core, now assumed to be constant throughout the cycle.

The inductance due to the core alone is then

$$L_m = L - L_a' = \frac{4N^2\mu_m A}{d} \times 10^{-9}, \quad \text{where} \quad L_a' = \frac{4N^2 A_a}{d} \times 10^{-9}. \tag{4}$$

The permeability of the core material can be obtained from this inductance as

$$\mu_m = \frac{L_m d}{4N^2 A} \times 10^9. \tag{5}$$

The peak flux density in the core is derived from eqs. (5) and (1) as

$$B_m = \mu_m H_m = \frac{\sqrt{2} L_m I}{NA} \times 10^8 \text{ gauss,} \tag{6}$$

where I is the r.m.s. current in the winding.

[3] A list of most frequently used symbols will be found in the appendix.

Simple Analysis of Eddy Current Resistance

Again, at bridge balance, the resistance of the standard is equated to the resistance of the test coil, which is composed of the copper resistance R_c and a resistance which corresponds to the a-c. power P dissipated in the core. Thus

$$R = R_c + P/I^2. \tag{7}$$

Power is dissipated in the core through eddy currents and magnetic hysteresis. Although both types of magnetic loss occur simultaneously, they will first be considered as if occurring alone, after which the details of separating and identifying the two types will be discussed.

The resistance due to eddy current power loss depends upon the form of the magnetic core—whether of laminations, wire, or powder—upon the frequency, upon the permeability of the magnetic material, and upon the hysteresis loss, since this modifies the permeability. It is determined with sufficient accuracy for many practical purposes by calculating the eddy current power loss in a volume element consisting of a thin tube so drawn that neither magnetic flux nor eddy currents cross its surfaces, when the flux it encloses varies sinusoidally, and then integrating between proper limits to include the entire cross-section of the lamination. By this method [4] it can be shown that the power consumption per unit volume of sheet core material is

$$P_1 = \frac{\pi^2 t^2 f^2 B_m^2}{6\rho} \times 10^{-7} \text{ watt}, \tag{8}$$

where t is the sheet thickness in cm., f is the frequency, and ρ is the resistivity of the material in e.m.u.

This relation is derived on the assumption of a very extensive plane sheet with magnetizing force parallel to its surface, but it applies sufficiently well to any sheet material, flat or curved, provided that the magnetizing force is parallel to its surface, and provided that the width of the magnetic sheet is large in comparison to its thickness. These conditions can be fulfilled in a core built up of ring shaped laminations, and wound with an annular winding.

The total eddy current power loss in a core of volume πAd is then

$$P_e = \frac{\pi^3 t^2 f^2 B_m^2 A d}{6\rho} \times 10^{-7}. \tag{9}$$

As already mentioned (eq. 7), such a power loss in the core of a coil

[4] C. P. Steinmetz, "A. C. Phenomena," p. 195 (1908).

appears in an a-c. bridge measurement as a resistance

$$R_e = \frac{P_e}{I^2} = \frac{\pi^3 t^2 f^2 B_m{}^2 A d}{6 \rho I^2} \times 10^{-7}. \tag{10}$$

Substituting for B_m from eq. (6), and for L_m from eq. (4), gives

$$R_e = \frac{4\pi^3 t^2}{3\rho} \mu_m L_m f^2 \text{ ohm.} \tag{11}$$

Since this equation contains explicitly no geometrical details of the core, other than the sheet thickness t, it is applicable to any type of core in which the flux density is uniform, as it is in an annular core. If the resistivity is expressed in microhm-cm., the eddy current resistance becomes

$$R_e = \frac{0.0413 t^2}{\rho_1} \mu_m L_m f^2 \text{ ohm.} \tag{12}$$

A similar solution for the case of a core consisting of a hank or bundle of insulated magnetic wires of diameter t cm. gives

$$R_e = \frac{\pi^3 t^2}{2\rho} \mu_m L_m f^2, \tag{13}$$

or with ρ in microhm-cm.,

$$R_e = \frac{0.0155 t}{\rho_1} \mu_m L_m f^2. \tag{14}$$

A compressed magnetic dust core can be idealized as composed of closely packed insulated spheres. Although there is considerable concentration of flux at various points in any practical core, the power loss in a sphere of diameter t_1 can be calculated to a first approximation by assuming it permeated by a uniform flux density parallel to the direction of the magnetizing force. Computing the eddy current power loss in a cylindrical shell of such a sphere with shell axis parallel to H, and integrating to obtain the total loss gives

$$P = \frac{\pi^3 t_1{}^5 f^2 B_m{}^2}{120 \rho} \times 10^{-7}. \tag{15}$$

The power expended in a cubic centimeter of such a core is then

$$P_1 = \frac{\pi^2 t^2 f^2 B_m{}^2 r}{20 \rho} \times 10^{-7}, \tag{16}$$

where t^2 is the mean square sphere diameter, and r is the packing

factor, i.e., the ratio of the volume occupied by metal to the total volume of the core.

The flux density in the spheres is larger than the apparent flux density by the factor $r^{-2/3}$. The eddy current resistance for such a structure then becomes

$$R_e = \frac{2\pi^3 t^2}{5\rho r^{\frac{1}{3}}} \mu_m L_m f^2, \tag{17}$$

where μ_m is the permeability of the core, as calculated from the inductance L_m.[5] With ρ in microhm-cm., this equation becomes

$$R_e = \frac{0.0124 t^2}{\rho_1 r^{\frac{1}{3}}} \mu_m L_m f^2. \tag{18}$$

In cases where merely comparative tests are to be made, or where ρ and t are not known, it is convenient to lump the coefficient of the eddy current resistance in the form

$$R_e = e\mu_m L_m f^2. \tag{19}$$

Simple Analysis of Hysteresis Resistance

In addition to the power loss due to eddy currents, there is a loss caused by magnetic hysteresis. The energy in ergs dissipated per cubic centimeter of core during one hysteresis cycle is

$$W = \frac{1}{4\pi} \oint H dB = \frac{a_1}{4\pi}, \tag{20}$$

where a_1 is the area of the hysteresis loop in gauss-œrsteds. The power consumption on this account in an annular core of volume πAd, carried through f cycles per second, is

$$P_h = W\pi dAf = \tfrac{1}{4} a_1 dAf \times 10^{-7} \text{ watt}. \tag{21}$$

In the same manner as above, this power is observed by an a-c. bridge balanced for the frequency f as a resistance

$$R_h = \frac{8\pi W}{B_m^{\ 2}} \mu_m L_m f = \frac{2a_1}{B_m^{\ 2}} \mu_m L_m f \text{ ohm}. \tag{22}$$

By this relation, the hysteresis resistance can be used to compute the energy loss per cycle W, or the hysteresis loop area a_1, which would be obtained by ballistic galvanometer measurements of sufficient sensitivity.

[5] Cf. R. Gans, *Phys. Zeit.* 24, 232 (1923).

The Rayleigh Hysteresis Loop

The hysteresis loop area for magnetic cycles at low flux densities (i.e. for which μ_m is not more than 10–20 per cent higher than μ_0) can be calculated from the general shape of such loops. Rayleigh found experimentally [6] that the two branches of such loops are parabolas, that the permeability corresponding to the tips of the loop increases in proportion to the peak magnetizing force, thus,

$$\mu_m = \mu_0 + \alpha H_m, \tag{23}$$

and that the remanent flux is

$$B_r = \frac{\alpha}{2} H_m^2. \tag{24}$$

The loop equation which satisfies the above conditions is

$$B = (\mu_0 + \alpha H_m)H \pm \frac{\alpha}{2}(H_m^2 - H^2), \tag{25}$$

where the points on the upper branch are obtained by using the $+$ sign and those on the lower branch by using the $-$ sign. Recent ballistic galvanometer measurements of high precision on an iron dust core tend to confirm the reliability of the Rayleigh loop equation for low flux densities.[7]

Integrating HdB around the cycle gives an area $4\alpha H_m^3/3$, which can be used in equation (22) to obtain the hysteresis resistance, as

$$R_h = \frac{8\alpha H_m}{3\mu_m} L_m f. \tag{26}$$

Defining the permeability variation with flux as

$$\lambda = \frac{\mu_m - \mu_0}{\mu_0 B_m}, \tag{27}$$

the value of α will be

$$\alpha = \mu_0 \mu_m \lambda, \tag{28}$$

and the hysteresis resistance becomes

$$R_h = \frac{8}{3} \lambda H_m \mu_0 L_m f. \tag{29}$$

[6] *Phil. Mag.* [5] **23**, 225 (1887).
[7] W. B. Ellwood, *Physics* **6**, 215 (1935).

It thus appears that the Rayleigh hysteresis loop implies a definite relationship between the variation of permeability with H or B, as calculated from bridge measurements of inductance at various coil currents, and the observed hysteresis resistance. Efforts to judge as to the general applicability of the Rayleigh form of loop by means of such a-c. bridge comparisons have indicated fairly good agreement for most materials, but occasional deviations as high as 40 per cent.[8] Best agreement is generally found in well annealed and unstressed materials, while deviations are found in such materials as compressed dust cores. In such comparisons, the anomalous residual loss, variously termed magnetic viscosity, and after-effect, is excluded. This additional loss will be discussed below.

Mutual Effect of Rayleigh Hysteresis and Eddy Current Shielding in Sheet Material

Taking the above equation as the simplest general representation of hysteresis loops at low flux densities, it now becomes necessary to review the previous work with additional refinements to include the effects of hysteresis upon eddy currents, and of eddy currents upon themselves, and upon hysteresis. Thus the fact that B varies according to a hysteresis loop equation rather than directly with H modifies the eddy current loss somewhat. Also, eddy currents set up magnetizing forces within the magnetic material which more or less neutralize that applied by the coil winding, and thus effectively shield the inner parts of magnetic laminations of wires. Such eddy current shielding reduces the total flux in the core, thus decreasing the inductance and loss resistance observed at higher frequencies.

The fundamental differential equation giving the relation between B and H at a point x distant from the median plane of a magnetic sheet is[9]

$$\frac{\partial B}{\partial t} = \frac{\rho}{4\pi} \frac{\partial^2 H}{\partial x^2} . \tag{30}$$

For the simple case of constant permeability in which $B = \mu_0 H$, this equation has been solved by Heaviside, J. J. Thomson,[10] and others.

For the case in which B is given by Rayleigh's equation (25), the solution is very much involved. The variable permeability gives rise

[8] E. Peterson, *B. S. T. J.* **7**, 775 (1928).
[9] *E.g.*, Russell, "Alternating Currents," Vol. I, p. 487 (1914).
[10] *Electrician* 28, 599 (1892).

to odd harmonic voltages which are important from the standpoint of modulation and noise. In a-c. bridge measurements where the balance is made so as to bring the voltages of fundamental frequency to equality, it suffices to carry through the mathematics for this frequency alone. This has been done for sheet and wire cores to an accuracy sufficient for most purposes by W. Cauer.[11] From his results for the inductance and power loss in a laminated core, the apparent permeability and loss resistance are calculated to be

$$\mu_{fm} = \frac{\mu_0}{\theta}\frac{\sinh\,\theta + \sin\,\theta}{\cosh\,\theta + \cos\,\theta} + \alpha H_m\left(1 - \frac{4\theta^2}{9\pi} - \frac{7\theta^4}{60} + \frac{2\theta^6}{45\pi} + \cdots\right) \tag{31}$$

$$= \mu_0\left(1 - \frac{\theta^4}{30} + \frac{\theta^8}{732} - \cdots\right) + \alpha H_m\left(1 - \frac{4\theta^2}{9\pi} - \frac{7\theta^4}{60} + \frac{2\theta^6}{45\pi}\cdots\right) \tag{32}$$

$$= \mu_0\left[1 + \lambda B_m - \frac{4\lambda B_m\theta^2}{9\pi} - \left(1 + \frac{7}{2}\lambda B_m\right)\frac{\theta^4}{30} + \frac{2\lambda B_m\theta^6}{45\pi}\cdots\right], \tag{33}$$

$$R_{fm} = \frac{2\pi f L_0}{\theta}\frac{\sinh\,\theta - \sin\,\theta}{\cosh\,\theta + \cos\,\theta} + \frac{8\alpha H_m f L_0}{3\mu_0}\left(1 + \frac{\pi\theta^2}{4} - \frac{7\theta^4}{60} - \frac{\pi\theta^6}{40} + \cdots\right) \tag{34}$$

$$= \frac{\pi f L_0\theta^2}{3}\left(1 - \frac{17\theta^4}{420} + \frac{\theta^8}{600} - \cdots\right) + \frac{8}{3}\lambda B_m L_0 f\left(1 + \frac{\pi\theta^2}{4} - \frac{7\theta^4}{60} - \frac{\pi\theta^6}{40}\cdots\right). \tag{35}$$

The quantity $\theta = 2\pi t\sqrt{\mu_0 f/\rho}$, where ρ is in e.m.u.; and $B_m = \mu_m H_m$, where μ_m is independent of f.

The hyperbolic function parts of these equations are valid at any frequency, but they give only those parts of μ and R which are due to the constant initial permeability μ_0. The series having α or λ as coefficients give the increases due to hysteresis.

The apparent permeability μ_{fm}, which is calculated from the measured inductance, decreases as the measuring frequency is increased. Furthermore, at higher frequencies, this permeability rises less rapidly with rise in measuring current than it does at low frequencies, and it will actually decline with increasing H at frequencies higher than that necessary to make $\theta > 1.6$, approximately. Thus, for the accurate determination of μ_0, μ_m, λ and α, it is necessary to make measurements at frequencies low enough to suppress these correction terms.

[11] W. Cauer, *Arch. f. Elektrotechnik* 15, 308 (1925).

The equations for resistance are similarly complicated. The first series gives that part of the eddy current resistance which is due to the constant μ_0. The coefficient of the second series indicates that this series involves the hysteresis resistance. However, terms in the second series which contain the factor f^2 will be recognized as eddy current components introduced by the fact that the permeability has been increased from the value μ_0 by the factor λB_m.

The complicated form and slow convergence of the above equation (35) for resistance make it difficult for use in interpreting a-c. bridge measurements. Considerable simplification is effected by dividing the observed resistance (eq. 35) by the observed inductance (from eq. 33) for each measuring current and frequency. Performing this operation, and rejecting series terms in λB_m higher than the first power, gives

$$\frac{R_{fm}}{L_{fm}} = \frac{4\pi^3 t^2}{3\rho}\mu_m f^2\left[1 - \frac{\theta^4}{140}(1 + 5\lambda B_m) + \cdots\right] + \frac{8}{3}\lambda H_m \mu_0 f\left[1 - \frac{\theta^4}{36}(1 - 5\lambda B_m) + \cdots\right]. \quad (36)$$

The coefficient of the first series is identical with the eddy current expression previously derived (eq. 11), which neglected eddy current shielding and hysteresis. The series itself, which includes these other effects, converges rapidly for $\theta < 1$, provided that the value of λB_m is not carried too high.

The coefficient of the second series is identical with the hysteresis expression derived from Rayleigh's equation (29) in which eddy currents were neglected. The second term of this series gives the amount by which eddy current shielding reduces hysteresis resistance at higher frequencies. It appears to converge less rapidly than the series for the eddy current resistance, but this is partly offset by the decrease of its second term with increase of λB_m. Thus, the coefficients of θ^4 become equal to 1/88 in both series if $\lambda B_m = 13/110$. This value of λB_m is reached when the flux density in the material is large enough to raise the permeability some 10 per cent above μ_0. Evidently, this value of λB_m can be exceeded somewhat without making the coefficients excessively large. However, if the measurements are made at too high flux densities, the hysteresis loops diverge more and more from the simple Rayleigh loop, and the present analysis becomes inapplicable.

In a-c. bridge measurements it is seldom desirable to measure at flux densities which will carry the permeability more than 10 per cent above its initial value. If measurements at higher flux densities are

desired, sufficient sensitivity can generally be obtained by wattmeter or ballistic galvanometer methods.

Graphical Separation of Losses

Since it is generally important to distinguish between types of magnetic losses, methods of analyzing the measurements have been devised to accord with the degree of refinement desired. A fairly simple graphical loss separation method is suitable if magnetic shielding can be ignored. However, it will be seen to lead to the inclusion of an additional term to account for the residual loss. If the effect of eddy current shielding is also to be considered, a more complicated analytical method of separation will be found necessary.

With magnetic shielding negligible, eq. (36) reduces to the form

$$\frac{R_m}{fL_m} = \frac{8}{3}\lambda H_m\mu_0 + \frac{4\pi^3t^2}{3\rho}\mu_m f \tag{37}$$

$$= \frac{8}{3}\lambda B_m\frac{\mu_0}{\mu_m} + \frac{4\pi^3t^2}{3\rho}\mu_m f \tag{38}$$

or

$$\frac{R_m}{\mu_m fL_m} = aB_m + ef. \tag{39}$$

The last form of the expression is most suitable for routine testing and design purposes. The hysteresis area constant a will be seen to be intimately related to the hysteresis loop area a_1 previously discussed; thus

$$a = \frac{2a_1}{B_m{}^3}. \tag{40}$$

Within the limits of applicability of Rayleigh's equation, the following relations also apply:

$$a = \frac{8\lambda\mu_0}{3\mu_m{}^2} = \frac{8\alpha}{3\mu_m{}^3}. \tag{41}$$

The losses observed on any test core can be separated graphically [12] by calculating the values of $R_m/\mu_m fL_m$ for a fixed value of H_m at all measuring frequencies, and plotting such values against frequency. The slope of the resulting straight line should then give e, and the intercept aB_m. When this process is repeated for other values of H_m, a series of intercepts will be obtained, all of which would be expected to yield a constant value for the loop area constant a. However, this is frequently found not to be true, but if the several intercepts so

[12] B. Speed and G. W. Elmen, *Trans. A. I. E. E.* 40, 596 (1921).

obtained be plotted against B_m, they generally fall upon a fairly straight line whose slope is a, and whose intercept on the line $B_m = 0$ is c. The value of a so obtained agrees fairly well in most cases with the value calculated from the permeability variation coefficient λ, which is the justification cited above for the Rayleigh equation. The presence of residual loss necessitates rewriting the loss equation with an additional term—

$$\frac{R_m}{\mu_m f L_m} = aB_m + c + ef. \tag{42}$$

The value of the intercept c, however, has no counterpart in the Rayleigh equation. It indicates the presence of a power loss proportional to the frequency, and thus similar to hysteresis, but contrarily proportional to the square of the magnetizing force, instead of to the cube. It is found not to contribute to harmonics or modulation generated by a core material, and might thus be represented by an elliptical increment to the Rayleigh loop.[13]

Residual loss has been ascribed to viscosity or "after-effect" in the core material.[14] The chief obstacle to this explanation is the observed constancy of c over a wide range of frequencies, in contrast to the variation to be expected from ordinary viscosity losses. Residual loss has been ascribed to inhomogeneities in the magnetic material [15] which lead to higher a-c. power losses than expected from the area of the hysteresis loop. This explanation seems promising, but the work to date has been chiefly qualitative, and it has not been shown to yield the required additional loss proportional to H^2. The parallel between this loss and eddy current loss, which is also proportional to H^2, is alluring, but the dependence of eddy current loss upon f^2 has remained a stumbling block. The mechanical dissipation of power through magnetostrictional motions seems also a possible explanation.[16]

Somewhat analogous to the residual loss is the excess eddy current loss generally observed. When the observed value of e is used to calculate the resistivity of a magnetic material, it generally gives a value somewhat smaller than the true resistivity, which indicates that the observed eddy current losses are correspondingly too large. The apparent resistivity so obtained approaches the true resistivity quite closely for well insulated laminations of pure, well annealed materials. It is interesting to note that the residual loss for such well

[13] H. Jordan, *Ann. d. Physik* [5] **21**, 405 (1934).
[14] H. Jordan, *E. N. T.* **1**, 7 (1924); F. Preisach, *Zeit. f. Phys.* **94**, 277 (1935).
[15] L. W. McKeehan and R. M. Bozorth, *Phys. Rev.* [2] **46**, 527 (1934).
[16] For a more thorough discussion see W. B. Ellwood, *Physics* **6**, 215 (1935).

annealed materials is also practically absent. For many materials, however, the apparent resistivity falls to 50–75 per cent of the true resistivity. The increase in eddy current loss thus observed is technically very undesirable since it necessitates rolling laminations considerably thinner than otherwise required, in order to suppress eddy current losses sufficiently.

The cause of extra eddy current losses in laminated material is definitely chargeable to the hard, low permeability surface of the material. The eddy current losses are determined largely by the high interior permeability, and the laminar thickness. The material near the surface conducts large eddy currents induced by interior material of high permeability, but it contributes very little to the average permeability for the entire sheet. Removal of low permeability surface material by etching [17] lowers the eddy current losses and increases the average permeability of the core, so that the apparent resistivity approaches more closely the true d-c. value. Of course, selection of material and proper mechanical working and heat treating technique are most desirable in avoiding at the outset such inhomogeneities, with their resulting excessive losses.

Analytical Separation of Losses

For special investigations where the accuracy of the graphical method of loss separation is not sufficient, it is necessary to return to the unabridged form of eq. (36), and employ an analytical method. For example with sheet material,

$$e = \frac{4\pi^2 t^2}{3\rho} \quad \text{and} \quad \theta^2 = \frac{3}{\pi} e\mu_0 f.$$

Rewriting eq. (36) with these substitutions and with an additional term to provide for the residual loss,

$$\frac{R_{fm}}{fL_{fm}} = aB_m\mu_m\left[1 - \frac{9e^2\mu_0^2 f^2}{36\pi^2}(1 - 5\lambda B_m) + \cdots\right] + c\mu_m + e\mu_m f\left[1 - \frac{9e^2\mu_0^2 f^2}{140\pi^2}(1 + 5\lambda B_m) + \cdots\right]. \quad (43)$$

It should be recalled that R_{fm} and L_{fm} are the core resistance and inductance measured at a definite current and frequency, while μ_m is the permeability of the core measured at the same current, but at a frequency low enough to make eddy current shielding negligible.

[17] Legg, Peterson and Wrathall, U. S. Patent 1,998,840 (1934).

Subtracting the value of R_{fm}/fL_{fm} for frequency f_1 from the corresponding value for frequency f_2, and dividing by the frequency interval $\Delta f = f_2 - f_1$, gives

$$\frac{\Delta\left(\dfrac{R_{fm}}{fL_{fm}}\right)}{\Delta f} = e\mu_m \left[1 - \frac{9e^2\mu_0^2}{140\pi^2}(1 + 5\lambda B_m)(f_1^2 + f_2^2 + f_1 f_2) - \frac{2}{3\pi^2} e\mu_0 \lambda B_m (1 - 7\lambda B_m)(f_1 + f_2)\cdots\right]. \quad (44)$$

An approximate value for e is sufficient in obtaining the correction terms in this equation. With the precise value of $e\mu_m$ thus obtained, the eddy current term in eq. (43) can be calculated for any frequency and permeability. Subtracting the proper eddy current term for each value of R_{fm}/fL_{fm} gives the hysteresis terms as remainders, which can be further analyzed in their relation to magnetizing force, as in the previous graphical loss separation. Loss separations, made thus precisely, reveal frequency variations of apparent resistivity and of the residual loss constant.

Capacitance, Leakance, and Eddy Current Loss of the Winding

In the discussion thus far, it has been assumed that the measured inductance and resistance of a test coil depend solely upon the core permeability and losses. This assumption must be modified under some conditions, for it is found that the distributed capacitance and leakance of the coil winding act as shunt impedances, which may diminish sufficiently at high frequencies to mask the actual inductance and resistance of the coil. Furthermore, the resistance of the test coil includes an amount corresponding to the power expended by eddy currents in the copper winding itself. It will be shown that such disturbing factors can generally be eliminated, either by modifications in the method of core loss separation, for materials in which eddy current shielding is negligible; or by winding the test core to give an inductance low enough to suppress such disturbing factors, for materials in which eddy current shielding is not negligible.

If the distributed capacitance and leakance can be considered as single lumps, C, and G, in parallel with the coil of inductance L and resistance R, the observed inductance at a frequency corresponding to $\omega = 2\pi f$ is found to be

$$L_{\text{obs.}} = \frac{L(1 - \omega^2 LC) - CR^2}{(1 - \omega^2 LC)^2 + 2GR + G^2(R^2 + \omega^2 L^2) + \omega^2 C^2 R^2}.$$

This simplifies at frequencies well below resonance to

$$L_{\text{obs.}} = L(1 + \omega^2 LC). \tag{45}$$

Thus the observed inductance tends to increase at higher frequencies on account of distributed capacitance, in contrast to its tendency to decrease on account of magnetic shielding in the core according to eq. (33). If the inductance L is known from low frequency measurements, and if computations from eq. (33) show that it does not decline appreciably because of eddy current shielding at the measuring frequency, the capacitance can be calculated from the relation

$$\omega^2 LC = \frac{L_{\text{obs.}} - L}{L}. \tag{46}$$

Similar complications arise in measuring the resistance of a coil at high frequencies. Under the same assumptions as above, the observed resistance is

$$R_{\text{obs.}} = \frac{R + G(R^2 + \omega^2 L^2)}{(1 - \omega^2 LC)^2 + 2GR + G^2(R^2 + \omega^2 L^2) + \omega^2 C^2 R^2}.$$

At moderate frequencies, this reduces to

$$R_{\text{obs.}} = (R + G\omega^2 L^2)(1 + 2\omega^2 LC), \tag{47}$$

from which it appears that leakance enters as an important part, and that the capacitance gives twice as large an increment for the resistance as for the inductance. The effect of distributed capacitance can be eliminated by dividing the observed value of resistance by $(1 + 2\omega^2 LC)$, where the correction factor is obtained from eq. (46). Thus

$$\frac{R_{\text{obs.}}}{1 + 2\dfrac{L_{\text{obs.}} - L}{L}} = R + G\omega^2 L^2. \tag{48}$$

The leakance term $G\omega^2 L^2$ can be eliminated as will be shown below.

The resistance term R includes the desired magnetic core resistance, but it also contains the resistance of the copper coil, which may have a considerable eddy current loss of its own. The copper eddy current loss occurs principally in the lower layers of the winding, which are cut by the alternating magnetic flux set up by the current in the winding. It is similar to the eddy current loss in the core material itself, varying with the square of the frequency, to a first approximation.[18] This loss must, therefore, be eliminated before accurate

[18] Cf. M. Wien, *Ann. d. Phys.* [4] **14**, 1 (1904); S. Butterworth, *Exp. Wireless* **6**, 13 (1929).

determination of the core loss is possible. The eddy current resistance of the copper winding is of the form

$$R_{ce} = e_c L_a f^2, \tag{49}$$

where L_a is the total air inductance of the winding, and e_c is the copper eddy current coefficient.

This eddy current coefficient is inversely proportional to the number of strands in the wire, so that it can be minimized by using wire consisting of many insulated strands. It increases somewhat with the number of layers in the winding. The coefficient may be determined for any type of winding by resistance measurements on an air core coil of dimensions and winding details similar to those of the magnetic core to be tested. Subtracting the eddy current resistance R_{ce} so computed, and the d-c. copper resistance R_c, from eq. (48) gives as the residual resistance

$$\Delta R = \frac{R_{\text{obs.}}}{1 + 2\dfrac{L_{\text{obs.}} - L}{L}} - R_c - e_c L_a f^2$$
$$= \mu_m L_m[(aB_m + c)f + ef^2] + G\omega^2 L^2. \tag{50}$$

This residual resistance consists of the core loss resistance, and an increment due to leakance. The latter can be largely suppressed by the use of low leakance insulating materials, by insuring that the winding is free from moisture, and by making the distributed capacitance as small as possible. Furthermore, it is known from experiments on the electrical conductance of insulating materials at elevated frequencies that the "quality" $Q = \omega C/G$ is practically a constant (C is the capacitance associated with G, — in this case the distributed capacitance). Inserting this value of G in eq. (50) gives

$$\Delta R = \mu_m L_m[(aB_m + c)f + ef^2] + 8\pi^3 CL^2 f^3/Q. \tag{51}$$

Theoretically, the coefficients in this equation can be obtained from resistance measurements taken at three different frequencies. Unavoidable errors in the measurements render such an analysis unreliable, so that it is generally preferable to obtain a larger number of observations, and to determine the coefficients graphically. Dividing by $\mu_m L_m f$, and neglecting the air inductance of the coil, the equation becomes

$$\frac{\Delta R}{\mu_m L_m f} = (aB_m + c) + ef + \frac{8\pi^3 CLf^2}{\mu_m Q}\,. \tag{52}$$

If the data at lower frequencies are sufficiently reliable, this parabola can be extrapolated to give the zero intercept $(aB_m + c)$ which contains the sought for hysteresis constants of the core. Subtracting the intercept so found, and dividing again by f gives

$$\frac{1}{f}\left\{\frac{\Delta R}{\mu_m L_m f} - (aB_m + c)\right\} = e + \frac{8\pi^3 CLf}{\mu_m Q}. \tag{53}$$

This is the equation of a straight line, when plotted against f. The intercept e is the desired eddy current coefficient for the core material. The slope of this line, S, yields the dielectric quality

$$Q = \frac{8\pi^3 CL}{\mu_m S}. \tag{54}$$

Here C is the distributed capacitance, which can be obtained from eq. (46). This relation is useful in comparing the qualities of various insulating and spacing materials, and in calculating the total losses to be expected in any proposed coil.

Accurate Separation by Limiting Inductance

It appears from the above discussion that magnetic loss separations can be made in spite of interference by distributed capacitance, leakance and eddy current resistance of the coil windings, provided that the interference is not too large, and provided that eddy current shielding in the test core is negligible. When the latter condition is not fulfilled, it becomes necessary to suppress the interference due to capacitance, etc., to negligibly small quantities. This is facilitated by proper technique in applying the windings, but any degree of suppression can be secured by sufficient limitation of the coil inductance, as will appear by reference to eq. (47). Although reduction of the coil inductance by using a winding with few turns is desirable in thus suppressing errors, it is undesirable in that it reduces the core loss resistance (cf. eq. 50) to a value which may be difficult to measure accurately on any available bridge. It is thus necessary to wind the test core to an inductance which will yield the largest possible loss resistance, without exceeding the allowable error from capacitance, leakance, and copper eddy current loss. The value of this maximum allowable inductance is obtained by calculating the inductance required to make the errors due to capacitance, leakance, and copper eddy currents at the highest measuring frequency equal to some tolerable small fraction of the core loss resistance.

A good separation of losses requires measurements at four or more frequencies, up to a point where the eddy current resistance is several times the hysteresis resistance, and at four or more values of measuring current in the useful range. The maximum frequency necessary to make the eddy current resistance mount to p times the hysteresis resistance is

$$f_m = \frac{ph}{e}, \qquad \text{where} \qquad h = (aB_m + c), \tag{55}$$

and the total core loss resistance at this frequency is

$$R_m = \frac{\mu_m L_m h^2 p(p+1)}{e}. \tag{56}$$

At this maximum frequency, the observed resistance is

$$R_{\text{obs.}} = (R_c + e_c L_a f^2 + R_m + 8\pi^3 C L^2 f_m^3/Q)(1 + 8\pi^2 L C f_m^2). \tag{57}$$

Since it is desired that the observed resistance indicate directly the d-c. copper and core loss resistance, all other terms in eq. (57) may be considered as errors, to be suppressed to a small fraction q of the core resistance. Setting the total error equal to qR_m, and rejecting errors of higher orders gives

$$qR_m = e_c L_a f_m^2 + 8\pi^2 L C f_m^2 (R_c + R_m + \pi L f_m/Q). \tag{58}$$

Substituting the above values for f_m and R_m at the maximum measuring frequency gives a quadratic equation in L (neglecting air inductance), which solves quite accurately to require,

$$L = \frac{e[eq\mu_m(p+1) - 8\pi^2 C R_c p]}{8\pi^2 C p^2 h[\pi/Q + \mu_m h(p+1)]} - \frac{e_c L_a p}{eq\mu_m(p+1) - 8\pi^2 C R_c p}. \tag{59}$$

Since it is desirable to use a large inductance for ease in resistance determination, it appears that the copper resistance R_c, the copper eddy current coefficient e_c, and the distributed capacitance C should be made as small as possible, while Q should be large.

As an illustrative example, assume a core material of permeability $\mu_m = 100$, to be measured up to a frequency such that $p = 5$, with an error of not more than 1 per cent at the maximum frequency, i.e., $q = 0.01$. Assume also, $e = 25 \times 10^{-9}$, $h = 0.5 \times 10^{-4}$, $C = 25 \times 10^{-12}$, $Q = 20$, $e_c L_a = 10^{-11}$, $R_c = 2$.

The maximum frequency for measurement will then be $f_m = 10{,}000 \sim$. The core should be wound to give an inductance of 5.30 mh. At the

maximum measuring frequency the eddy current resistance will be 1.325 ohms and the hysteresis resistance 0.265 ohm. At the lower end of the frequency range, say at 1000 ~ these figures become $R_e = 0.0132$ ohm and $R_h = 0.0265$ ohm. This indicates that a bridge will be required for such measurements capable of measuring increments of resistance to an accuracy of about 0.0002 ohm at 1000 ~, and about 0.002 at 10,000 ~.

Computations of this sort show the high quality of a-c. bridge generally demanded for core loss measurements. Such measurements require equipment and a bridge with a maximum of sensitivity of both a-c. and d-c. balances, and a minimum of losses in standards, pick-up, unbalanced impedance to ground, and variable contact.

Bridges, and Test Procedure

The essential features of bridges suitable for core loss measurements will now be mentioned in general terms. It will be appreciated from the above discussion that the specific range of frequencies, and the required resistance sensitivity must be adapted to the loss characteristics of the magnetic core to be measured.

Although certain types of resonance bridges [19] have advantages for measurements at high frequencies, the most suitable bridge for the usual measurement of magnetic core coils is an equal arm inductance comparison bridge,[20] on which inductances can be measured directly, and on which a-c. and d-c. resistance measurements can be made in prompt succession, to eliminate the effect of gradual temperature changes on the resistances of the bridge and test coil. A suitable circuit is shown in Fig. 1.

Inductance coils for bridge standards should be as stable as possible against frequency and current. Although low effective resistance per unit of inductance is desirable, it is more important for core loss measurements to design such standards for a minimum increase of resistance with frequency, so as to keep calibration corrections small in comparison to the resistance increments to be measured. A satisfactory type of standard coil consists of an air core toroidal form, with a bank winding of finely stranded wire. The bank winding minimizes capacitance effects on the observed inductance and resistance of the coil, and the fine stranding minimizes eddy current losses in the copper; cf. eq. (49). The small residual corrections must finally be included as calibrations when making measurements with the aid of standard coils.

[19] W. J. Shackelton and J. G. Ferguson, *B. S. T. J.* **7**, 82 (1928).
[20] W. J. Shackelton, *B. S. T. J.* **6**, 142 (1927).

The effect of contact resistance is minimized by changing few, or preferably no, contacts between a-c. and d-c. readings. This is facilitated by supplying current to one corner of the bridge through the sliding contact of the slide wire resistance used for fine balancing. This excludes contact resistance errors from resistance determinations in which both a-c. and d-c. balances fall within the range of the slide wire, and thus increases the bridge accuracy for small values of effective resistance. Usual precautions as to clean and positive contacts are sufficient for larger resistance measurements.

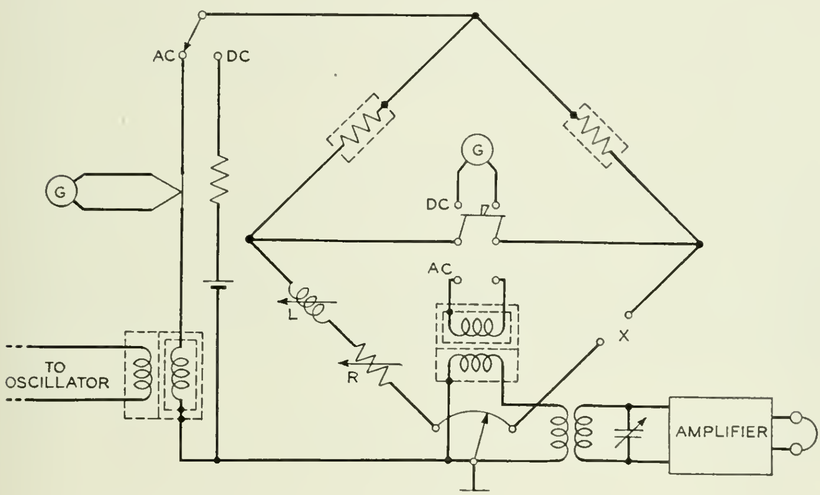


Fig. 1—Diagram of inductance comparison bridge suitable for measurement of magnetic core coils.

The a-c. supply to the bridge should, of course, be a sine wave, and the bridge transformers should be designed for minimum distortion. The frequency should be known accurately, and the voltage should be constant during any set of measurements. Rheostats are required to permit accurate adjustment of bridge current, and they must be designed and shielded to avoid stray coupling with the bridge. A suitable thermocouple is provided for measuring the current into the bridge, from which measurement the current through the test coil can be readily determined, since the bridge has equal ratio arms.

A distortion-free amplifier and a filter circuit tuned to the measuring frequency, are essential for magnifying the bridge unbalance current so as to permit precise measurements. Such unbalances may be detected by a vibration galvanometer at frequencies below say 200 ∼

and by head phones at frequencies within the audible range. For measurements at higher frequencies, a heterodyne detector is needed. Detection with a galvanometer is feasible when used in connection with a rectifier and filter, the filter being required so as to eliminate errors due to currents of extraneous frequencies.

The galvanometer used for d-c. bridge balancing should be sensitive enough to secure resistance readings of precision equal to that of the a-c. balance. The d-c. supply to the bridge should be limited to a current of the same order of magnitude as the a-c. supply, to guard against permanent magnetization of the magnetic material under test.

The usual test procedure is to set the oscillator at the lowest desired frequency, and measure the inductance and resistance of the test coil at several currents beginning at the lowest, increasing to the highest, and returning again to the lowest, so as to detect any tendency for permanent magnetization or magnetic aging. Direct current balances are taken as often as required to keep up with gradual changes of circuit resistance due to room temperature changes, the direction of current through the bridge being reversed each time to detect and eliminate stray currents and thermal e.m.f.'s. The differences between the observed a-c. and d-c. resistances gives the a-c. increment resistance of the test coil, except for corrections on account of the calibration of the bridge and coils. This process is repeated at successively higher frequencies until a suitable range of data has been covered. The resulting data can then be analyzed to show the characteristic of the magnetic core by the appropriate method as described above.

The Test Core

The design of the test core depends upon the physical and magnetic characteristics of the material to be tested. In general the radial thickness should be small in comparison with the diameter. Strain sensitive materials must be protected from mechanical stresses of handling and winding. Types of insulation and winding depend upon the loss characteristics of the core.

In any practical core, the diameter ranges between an inside value d_i and an outside value d_0. Since the diameter enters into the denominator of the expression for H and certain other magnetic quantities, the effective diameter must be calculated and used in such expression rather than the simple mean diameter. The effective magnetic diameter of a core having a rectangular cross-section is

the reciprocal of the value obtained by averaging $1/d$, namely

$$d = \frac{d_0 - d_i}{\log_e \dfrac{d_0}{d_i}}. \tag{60}$$

This expression can be converted to the following convenient series

$$d = d_m \left[1 - \frac{1}{12} \left(\frac{\Delta d}{d_m} \right)^2 - \frac{1}{180} \left(\frac{\Delta d}{d_m} \right)^4 - \cdots \right], \tag{61}$$

which indicates that the effective diameter is smaller than the arithmetical mean diameter d_m, by an amount depending upon Δd, the difference between inside and outside diameters. This series converges so rapidly that terms beyond the second may be neglected for all practical purposes.

A test core of large radial thickness is to be avoided, when accurate measurements are desired, because of the considerable change of flux density from the inside to the outside diameter, with its accompanying modification of the core permeability. Such variations complicate eddy current and hysteresis behavior, particularly through reaction on the magnetic permeability. If the permeability at every point in the core bears a straight line relationship to the flux density, eq. (27) gives

$$\mu = \mu_0(1 + \lambda B). \tag{62}$$

The flux density at diameter y is

$$B = \mu H = \frac{0.4 N i \mu_0}{y} (1 + \lambda B). \tag{63}$$

Solving this equation for B, and integrating from d_i to d_0 gives the total flux in a unit height of core, from which the mean flux density can be calculated. This gives for the mean permeability, approximately

$$\mu_m \doteq \mu_0 \left(1 + \lambda B_m \frac{d^2}{d_i d_0} \right), \tag{64}$$

where B_m is the mean peak flux density in the core, and d is the effective magnetic diameter. Comparison of (64) with (62) shows that the increase of mean permeability in a core of considerable radial thickness is not precisely equal to the ideal increase of permeability for a given flux density.

The effect of radial thickness of core on losses can be attacked in a similar manner. Inserting the value of H at diameter y in Cauer's

expression for power loss,[21] computing the loss in a ring of thickness $dy/2$, and integrating from d_i to d_0 yields an expression for core resistance identical with (33), (34) and (35) except that the terms containing α or λ must be multiplied by the factor $d^2/d_i d_0$. It appears that radial thickness of the core affects only that part of the loss which depends upon the variation of permeability with H or B.

In preparing a test core, a compromise must be struck between the radial thickness, diameter, and axial height, so as to secure the desired cross-sectional area without excessive length of copper winding. The laminations of the core must be well insulated from each other. They should be of uniform thickness, which should be known accurately. A good technique used with material in the form of ribbon or tape, consists of winding it tightly in several layers upon a cylindrical mandrel and providing insulation against eddy current straying by dusting the strip with finely powdered alumina while winding. Such insulation is found to withstand the high temperatures ordinarily used in heat treating the core. In order to eliminate the airgap in this type of core, the inside end of the tape may be brought out, folded over, and welded to the outside end, before annealing the core. The use of 50 turns of tape or more in a spiral core reduces the airgap effect to a negligible amount so that welding the tape ends is not necessary. A mandrel diameter should be selected large enough to make the ratio $\Delta d/d_m$ quite small. Thus, a 9 cm. mandrel, wound to a depth of 1 cm., gives a core in which the magnetizing force decreases about 20 per cent from the inside diameter to the outside, while the correction term decreases the effective diameter about 0.4 per cent below the mean diameter. The correction to the permeability variation term is $d^2/d_i d_0 = 1.002$.

The completed core, if strain sensitive, can be protected from the mechanical stresses of handling and winding by mounting it in a loose fitting toroidal box upon which the test windings are applied. Such spacing helps to decrease distributed capacitance, but even more important is sectionalizing, or bank winding of the coil. Types of insulation and windings depend upon the loss characteristics of the core. In general, lower loss characteristics in the core require higher quality windings, to permit measurements at higher frequencies.

SYMBOLS

a Hysteresis resistance coefficient.

a_1 Hysteresis loop area; $= \frac{1}{2}aB_m{}^3 = 4\pi W$.

[21] W. Cauer, *Arch. f. Elektrotechnik* **15**, 308 (1925).

A Cross-sectional area of magnetic core; cm.^{2}

α Permeability-magnetizing force coefficient; $= \mu_0\mu_m\lambda$.

B Instantaneous flux density in core; gauss.

B_m Maximum flux density in core subject to alternating magnetizing force.

c Residual resistance coefficient.

C Distributed capacitance of coil winding; farad.

d Effective magnetic diameter of annular core; cm.

e Eddy current resistance coefficient of core.

e_c Eddy current resistance coefficient of coil winding.

f Frequency of alternating current.

G Distributed leakance of coil winding; mho.

h $= (aB_m + c)$.

H Instantaneous magnetizing force in core; œrsted.

H_m Maximum magnetizing force in core subject to alternating magnetizing force.

i_m Maximum of alternating current wave in coil; ampere.

I Effective or r.m.s. current in coil.

L Inductance, due to core and residual air space only; henry.

L_a' Inductance due to residual air space.

L_a Inductance due to coil with air core.

L_m Inductance with current of maximum value i_m due to core only.

L_{fm} Inductance due to core only, with current of maximum value i_m, at frequency f (i.e., subject to magnetic shielding).

$L_{obs.}$ Total inductance observed at frequency high enough to give increases due to distributed capacitance and leakance of the coil winding.

λ Permeability-flux density coefficient; $= \dfrac{\mu_m - \mu_0}{\mu_0 B_m}$.

μ_0 Initial permeability of core.

μ_m Permeability corresponding to L_m.

μ_{fm} Permeability corresponding to L_{fm}.

p Ratio of eddy current to hysteresis resistance at maximum frequency.

P Total power dissipated in core; watt.

P_e Eddy current power dissipated in core.

P_h Hysteresis power dissipated in core.

q Fraction of core loss tolerated as error due to capacitance, leakance, and copper eddy current loss in winding.

Q Insulation quality factor $= \omega C/G$.

R Resistance due to core and winding only; ohm.

R_c Direct current resistance of copper winding.

R_{ce} Eddy current resistance of copper winding.
R_e Eddy current resistance due to core.
R_h Hysteresis resistance due to core.
R_{fm} Resistance due to core only, corresponding to L_{fm}.
$R_{\text{obs.}}$ Total resistance corresponding to $L_{\text{obs.}}$.
ΔR Resistance due to core and leakance only.
ρ Resistivity of core material; abohm-cm.
ρ_1 Ditto; microhm-cm.
t Thickness of sheet, diameter of wire, or r.m.s. sphere diameter of magnetic material; cm.
θ Eddy current parameter; $= 2\pi t\sqrt{\mu_0 f/\rho}$ for sheet material.
W Hysteresis energy dissipated per cycle per cubic centimeter of core; erg.

The Present Status of Ferromagnetic Theory *

By R. M. BOZORTH

DISCOVERY of the loadstone and some of its magnetic properties is now reputed to be some three thousand years old. During these many years ferromagnetism has resisted very successfully the attack of theorists, and even at the present time theory lags far behind experiment. But advances in theory have been particularly rapid during the last five or ten years; the author describes in this paper what he regards as the high points of this progress.

Not until the last quarter of the last century was any considerable work done on magnetic materials. During this period data were gathered rapidly until, just before the close of the century, an excellent book [2] of four hundred pages, containing practically all of the important experimental and theoretical facts, was written by J. A. Ewing, later Sir James Ewing. The shape of the magnetization curves of iron, cobalt, and nickel, the existence of magnetic saturation and the magnetic transformation temperature, the existence and some of the laws of hysteresis, the simpler effects of stress and of magnetostriction, together with the important methods of measurement—all were known then, and silicon steel had just been invented.

Strangely enough, during the next fifteen years there was but little advance in knowledge of magnetic materials, but there were many applications of existing knowledge by engineers to electrical machinery, including those in electrical communication. During this period, also, the Heusler alloys (non-ferrous alloys exhibiting ferromagnetic properties) were invented; and although these served to stimulate those interested in the theoretical aspects of ferromagnetism, still there was little progress.

Beginning between 1915 and 1920 and extending to the present, there has been a rapid development on both the experimental and theoretical sides of ferromagnetism. To illustrate the progress that has been made in the improvement of magnetic materials, Table I has been prepared. The improvements made during the last 20 years have resulted from new methods of purification of the materials, new

* This paper as herein published contains a few revisions and additions to the paper as it appeared in the November 1935 issue of *Electrical Engineering*. It is scheduled for presentation at the A. I. E. E. Winter Convention, New York, N. Y., January 28–31, 1936. A subsequent paper in the same field of endeavor is planned in which entirely new material will be presented.

TABLE I

Some Extremes in the Properties of Magnetic Materials Available in 1915, and in 1935

Material	Property	Value 1915	Value 1935
Iron	Maximum permeability	45,000[11]	340,000[12]
	Initial permeability	300	20,000[12]
	Coercive force in œrsteds	0.3[11]	0.03[12]
Iron-nickel [13]	Maximum permeability	2,800[14]	600,000[15]
	Initial permeability	700[16]	12,000[17]
	Coercive force in œrsteds	1.5[14]	0.01[15]
Silicon-iron	Initial permeability	400	2,000[12]
Iron	Hysteresis at $B_M = 100$ gausses, in ergs per cm.^{3} per cycle	20	0.1[12]
Iron-cobalt-nickel "perminvar"	Hysteresis at $B_M = 100$ gausses, in ergs per cm.^{3} per cycle	....	0.00003[18]
Iron-cobalt	Saturation value in gausses [19]	25,800	25,800
	Permeability at $B = 16,000$ gausses	2,100	19,000[18, 12]
Tungsten steel	Coercive force [4] in œrsteds	80	80
New K. S. steel	Coercive force [21] in œrsteds	....	900

Superior numerals refer to references at end of paper.

compositions (alloys), and new methods of heat treatment. Some of these figures refer only to laboratory specimens, and not to materials available in commercial quantities.

But the chief topic of this paper is the theoretical side of ferromagnetism. How is one to explain the different values of magnetic permeability, ranging from 1 to 600,000 for various materials? Or, to consider first the more fundamental questions, what is the elementary magnetic particle, and why is ferromagnetism associated with so few elements?

Origin of Ferromagnetism

It was suggested by Ampère about one hundred years ago that molecules might behave as magnets because of the electric currents circulating in them. Today, with the advance in knowledge of atomic structure, the origin of ferromagnetism can be discussed in more specific terms. Strangely enough, the spectroscopists have supplied, so to speak, the elementary magnetic particle. It is the spinning electron. In order to explain their extensive observations on spectral lines, they found it necessary to revise the picture of the atom. For some time it has been supposed that an atom was made of a heavy nucleus with a positive charge and of electrons moving in circular

or elliptical orbits around the nucleus. To this picture now must be added the idea that each electron itself is spinning about an axis that passes through its center. Thus, there is circulation of electricity in an atom, both around the nucleus and within each electron—and the latter motion is called the "electron spin" because of its similarity to a spinning ball. Each electron in an atom is then a small gyroscope, possessing a definite magnetic moment on account of its moving electrical charge and a definite angular momentum on account of its moving mass. The ratio of these two quantities is known from various independent lines of reasoning and evidence to possess a particular value. Electrons revolving *in orbits* also exhibit both magnetic moments and angular momenta due to their orbital motions, but for these the ratio is just half what it is for the spinning electron.

The Barnett experiment [22] shows in a very direct way the existence of these magnetic and mechanical moments of the electron and confirms the ratio between them in ferromagnetic materials (Fig. 1). A

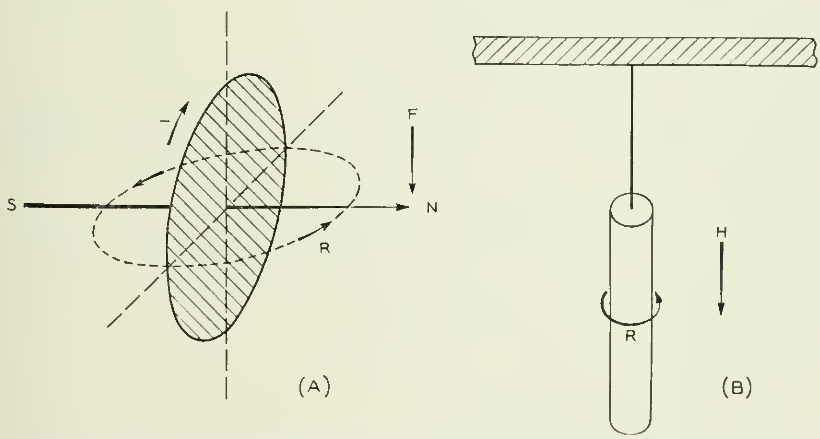
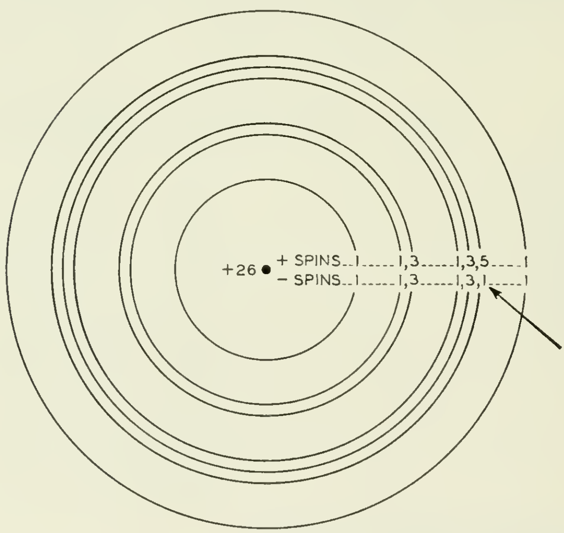


Fig. 1—Gyroscopic action (left); force F produces rotation R. Gyromagnetic effect (right); field H produces rotation R.

rod of iron is hung from a fine suspension and then is magnetized suddenly, whereupon the rod is observed to turn, twisting the suspending fiber a minute but measurable amount. The spinning electrons responsible for ferromagnetism have been turned by the applied field so that they are more nearly parallel to it; but the mechanical moment, which is also a property of those same electrons, causes the whole rod to rotate in just the way that a gyroscope would. Or, to put it differently: When the elementary magnets, pointing originally in all directions, are turned more nearly into parallelism with the axis of

the rod by the applied field, they acquire a net angular momentum parallel to that axis. By the principle that action must be balanced by a corresponding reaction, the rod itself now must recoil with an equal and opposite momentum; it is this last that manifests itself by the sudden twist of the rod and may be calculated from the measured value of the twist. Its sign shows that the spinning magnetic particle is charged negatively, and its magnitude is what would be expected from the hypothesis that that particle is a spinning electron. Thus a change in magnetization is fundamentally a change in the direction of the spin of the electrons in the atom, and not a change in orientation of the whole electron orbit.

The next question is: Why is not every substance ferromagnetic? The picture of the atom of iron as now envisioned by the experts in this field, is represented by the diagram in Fig. 2. The twenty-six elec-



	+ SPINS	− SPINS	EXCESS SPINS
Cr	4	0	4
Mn	5	0	5
Fe	5	1	4
Co	5	2	3
Ni	5	3	2
Cu	5	5	0

Fig. 2—Electron shells in the iron atom.

trons in iron are divided into four principal "shells," each shell a more or less well defined region in which the electrons move in their orbits, and some of these shells are subdivided. The first (inner) shell contains two electrons, the next shell eight, the next fourteen, and the last two. As the periodic system of the elements is built up from the lightest element, hydrogen, the formation of the innermost shells begins first. When completed, the number of electrons in the first four shells are two, eight, eighteen, and thirty-two, counting outward,

but the *maximum* number in each shell is not always reached before the next shell begins to be formed. For example, when formation of the fourth shell begins, the third shell contains only eight electrons instead of eighteen; it is the subsequent building up of this third shell that is intimately connected with ferromagnetism. In this shell some electrons will be spinning in one direction and others in the opposite, and these two senses of the spins may be conveniently referred to as positive and negative. The numbers inserted in Fig. 2 show how many electrons are present in each shell which have positive and negative spins, and it may be noticed that in the iron atom all of the shells except the third contain as many electrons spinning in one direction as in the opposite. The magnetic moments of the electrons in each of these shells mutually compensate one another so that the shell is magnetically neutral and cannot have magnetic polarization. In the third shell, however, which is not yet filled to this extent, there are five electrons with a positive spin and one with a negative so that four electron spins are (unbalanced or) uncompensated and there is a resultant polarization of the atom as a whole. If one more positive charge and its associated mass (a proton) be added to the nucleus and one more electron to an outer shell, the iron is transformed into cobalt; and by repeating the process, the cobalt is transformed into nickel. In iron these additional electrons and their spins are so oriented that there is what may be called an excess spin of four units in iron, three in cobalt, and two in nickel. In manganese, the element just preceding iron in the periodic table, there is an excess of five spins. Only in incomplete shells such as this, shells that are being filled as new and heavier atoms are made, is there such excess spin. The completed shells are magnetically neutral because the spins mutually compensate one another.

The outermost electrons are those responsible for the ordinary chemical properties, and they are influenced by chemical combination. They do not contribute to ferromagnetism for reasons that will appear later.

Exchange Forces

Only in certain parts of the periodic table are there found electrons being added to *inner* shells, and one of these places is in the iron group; but since there are other parts, notably those occupied by the palladium, platinum, and rare earth metals, where these inner groups are being filled, there arises the further question: Why are not these other elements also ferromagnetic?

For an element to be ferromagnetic, it is necessary not only that there be uncompensated spin in the electron orbits, but also that the

resultant spins in neighboring atoms be parallel. Calculation of the energies of the electrons indicates that to align the spins in all the atoms in a small region, the diameter of an atom must bear the proper ratio to the diameter of the electron shell in which the electron spins are uncompensated [23] (Fig. 3). This proper ratio is required because

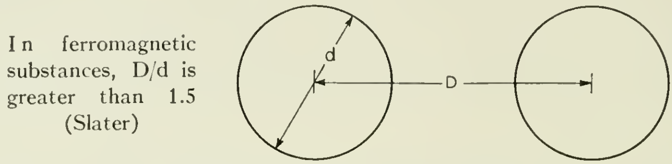


Fig. 3—"Incomplete" shells in neighboring atoms.

the electron spins and charges influence each other to an amount depending upon the distance between them; and it is only when this influence, which is known technically as the "exchange," has the right value that the spins all can be aligned in the same direction,[24] that is, that the material can become ferromagnetic.[25]

The forces of "exchange," the existence of which has been realized only in the last few years, act to keep the spins parallel, while thermal agitation tends, obviously, to disturb this alignment. When the temperature is high enough, the temperature agitation prevails and the material ceases to be ferromagnetic. This temperature is the familiar Curie point, or magnetic transformation point, 770 degrees centigrade (or 1,043 degrees absolute) for iron. It is seen then that the height of the Curie point, θ (on the absolute temperature scale) is an indication of the strength of the forces of exchange, which cannot yet be calculated theoretically except as to order of magnitude. These Curie points are plotted in Fig. 4 for the elements near iron in the periodic table; if a continuous curve be drawn through the points, it has a maximum near cobalt. Now the saturation value of magnetization depends both on the exchange and on the number of effective electron spins, that is, upon the number of electrons that can be oriented parallel to the field and the strength of the forces that hold them parallel. In a very rough way, it may be said to depend on the product of the exchange and the number, S, of uncompensated spins in the atom. Adopting θ as a measure of the exchange forces and forming the product θS, the right-hand curve in Fig. 4 is obtained, which indicates that the highest saturation should be attained in an iron-cobalt alloy, and that under certain appropriate conditions manganese might be ferromagnetic. Both these indications are substantiated by the data: The only known alloys having a higher

saturation value than pure iron are the iron-cobalt alloys; and compounds and alloys of manganese are more magnetic than any others that do not contain iron, cobalt, or nickel. The Heusler alloys, composed of manganese, aluminum, and copper, have a saturation almost as high as nickel, and numerous compounds of manganese are ferromagnetic in a less degree.

The forces of exchange are purely electrostatic in origin. But they are not electrostatic in the classical sense of the word; they are the result of electric charges distributed in space in a definite way. It does not seem to be possible to describe them easily in words, for it takes a great many mathematical equations to derive the result, which is a consequence of the assumptions of quantum mechanics. These forces

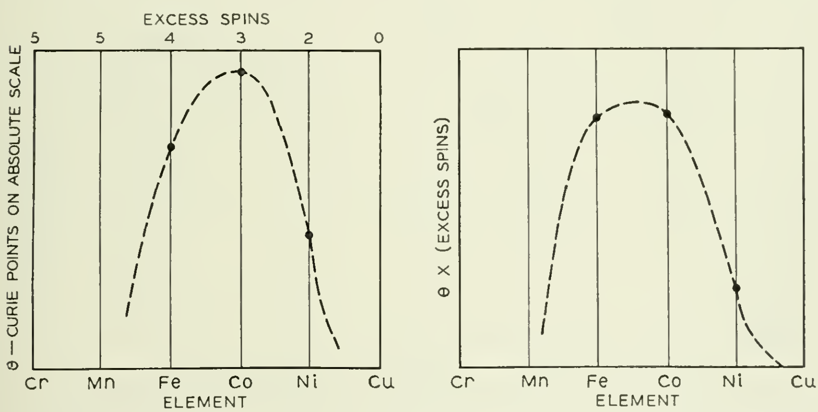


Fig. 4—Iron-cobalt alloys have the highest value of saturation magnetization.

account for the fact that it is easy to align the excess spins of large groups of atoms of some materials. In fact, when the forces of exchange are large as they are in ferromagnetic materials, the stable situation is one in which the spins are parallel, even when no magnetic field is applied. But the parallelism under such circumstances does not extend over the whole of a specimen of ordinary or even of visible size; for some reason not understood it is limited to smaller regions. On the average, those regions are found experimentally to have the volume of a cube about 0.001 inch on an edge. *An actual ferromagnetic body is composed of a great many such regions, called "domains," each domain being magnetized to saturation (i.e., electron spins parallel) in some direction.* When the material is said to be unmagnetized, the domains are oriented equally in all directions so that the magnetization of the specimen as a whole is zero.

Experimental evidence of the existence of these domains is supplied by the so-called "Barkhausen effect" (Fig. 5). If a small portion of a magnetization curve such as is shown in Fig. 5 could be magnified a billion times, it would be seen to be made up of steps, each a sudden change in magnetization as the field is increased, with no further change until the field reaches a certain higher value. No known apparatus can give such direct magnification, but these sudden jumps can be detected by winding a coil around the specimen and connecting its ends to an amplifier at the output of which is a pair of telephone

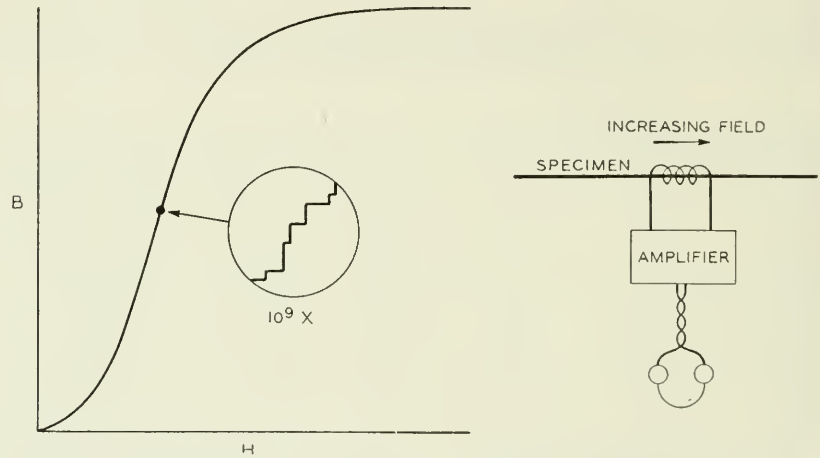


Fig. 5—Sudden changes in magnetization cause the Barkhausen effect.

receivers. When the field is slowly increased, a series of clicks, or "noise," is heard in the receivers; a more quantitative method shows that the average click corresponds to the reversal of magnetization in a region the size [26] of a cube 0.001 inch on an edge, containing 10^{15} atoms. Under favorable conditions this "Barkhausen noise" can be heard without an amplifier, with the receivers connected directly to the coil.

It has been pointed out that the forces of exchange are opposed by the disordering forces of temperature agitation. As a result, the saturation value of magnetization decreases continuously as the temperature is increased, until at the Curie point the ferromagnetism disappears. Data for saturation at various temperatures are shown in Fig. 6, plotted in such units that the saturation is unity at the absolute zero of temperature, and the Curie point is unity on the temperature axis. On such a plot it is found that the data for iron, cobalt, and nickel fall close together. The lower curve is the theoretical one calculated

thirty years ago on the assumption that the elementary magnets, when they are disturbed by temperature agitation, can assume any orientation. If it be assumed, on the contrary, that the spinning electrons responsible for ferromagnetism can assume only two orientations with respect to the other electrons in the atom, the upper curve is the result. If we assume that four orientations are possible the calculated curve lies close to the upper curve of Fig. 6, but somewhat

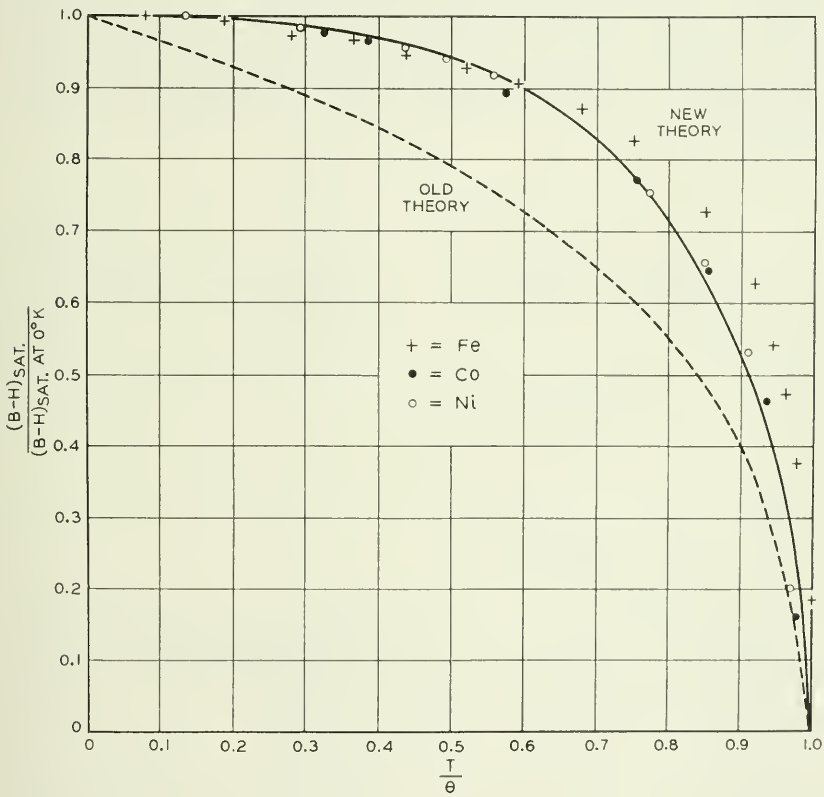


Fig. 6—Dependence on temperature of the saturation magnetization of iron, cobalt, and nickel.

below it, and as the number of possible orientations is increased the curve approaches the lower one shown in the same figure. The close agreement between the data and the upper calculated curve is of special interest because the spectroscopists and atomic structure experts have come independently, each group from its own data, to the conclusion that each electron in an atom can assume only a small number of orientations with respect to the rest of the atom.

Effect of Crystal Structure

There is another kind of force that must be postulated in order to explain the properties of a single crystal. Because of the spinning electrons which it contains, and also because of their orbital motions, each atom may be regarded as a small magnet. These magnets will influence each other in a purely magnetic way,[27] just as a group of bar magnets will; and in a crystal it may be readily appreciated that because of these magnetic forces between atoms arranged in a regular fashion, some directions of magnetization are more stable than others. In iron the most stable direction is observed to be that of the cube edge, one of the cubic axes of the crystal. In nickel it is the cube diagonal (Fig. 7).

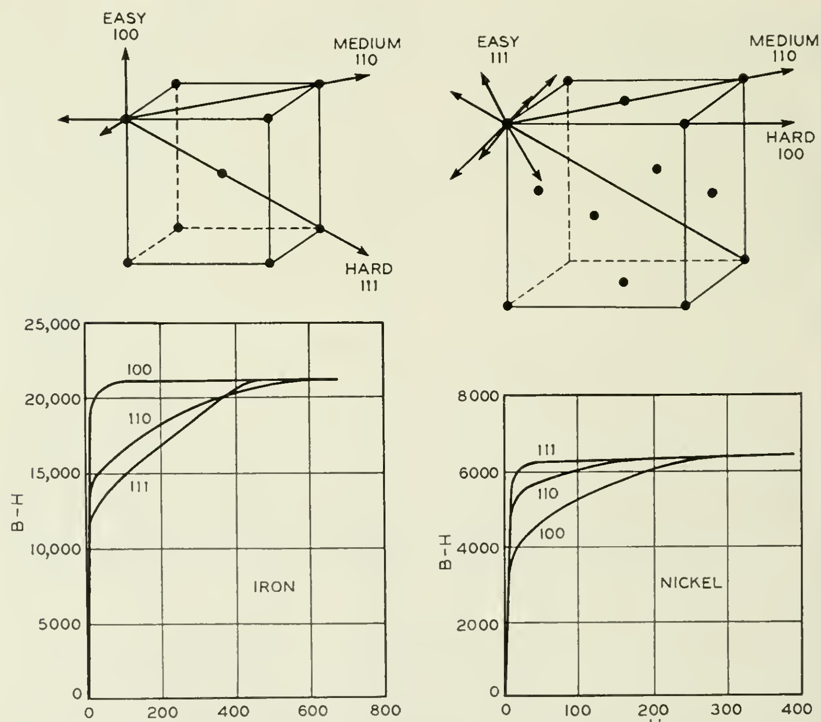


Fig. 7—Magnetic properties and crystal structures of single crystals of iron and nickel (Beck, Honda and Kaya, Webster).

Ordinarily a piece of iron is composed of crystal grains each one of which is too small to be detected by the naked eye. In recent years, however, means have been found to control the grain size of all the

common metals, and single grains (i.e., single crystals) have been prepared which are so large that experiments may be performed and data collected on just one such crystal.

The structure of a single crystal of iron may be represented by a cube with an atom at each corner and one in the center, the whole crystal made up of such cubes packed together face to face. It is found experimentally that in the direction of an edge of this cube (called by the crystallographers a [100] direction) the magnetization curve labeled 100 in Fig. 7 is obtained.[28] In the two other principal directions, the direction of a face diagonal and that of a cube diagonal, the other magnetization curves are obtained, as shown. The difference in the initial parts of the magnetization curves is negligible, the effects being large only above half saturation.

The structure of nickel may be represented also by an assemblage of cubes, but the atoms are arranged in a different manner, being at the corners of the cubes and the centers of the cube faces (Fig. 7). The magnetization curves for nickel corresponding to the same three principal directions are shown also in Fig. 7, and it may be seen that the curves are reversed in order from those of iron. In iron the [100] direction is said to be the direction of easy magnetization and the [111] the direction of most difficult magnetization, whereas the reverse is true in nickel. It might be said that the electrostatic exchange forces align the spins parallel to each other and that the crystal forces determine the particular crystal direction along which they shall be aligned. The forces of exchange are so powerful that they are able to align the spins of a group of atoms, a situation that in the absence of such exchange forces could be accomplished at room temperature only by an applied field of 10,000,000 œrsteds. On the other hand, the crystal forces are so feeble that it takes only 1,000 œrsteds to redirect the spins of an entire group of atoms from any direction to any other direction. The ratio between these two equivalent fields is thus 10^7 divided by 10^3, or 10^4.

As a result of the forces of exchange and the magnetic crystal forces in a single crystal of iron, for example, the situation is as represented in Fig. 8. Even when the crystal is apparently unmagnetized, or demagnetized, there are small regions, called domains, that are magnetized to saturation in one of the six equivalent directions of the crystal axes. Actually, the domains vary considerably in size and shape, but are represented conveniently as squares. Each of the six directions is equally stable and equally probable when no field is applied. The initial effect of applying a magnetic field is to change the direction of magnetization from one stable position to another, thus

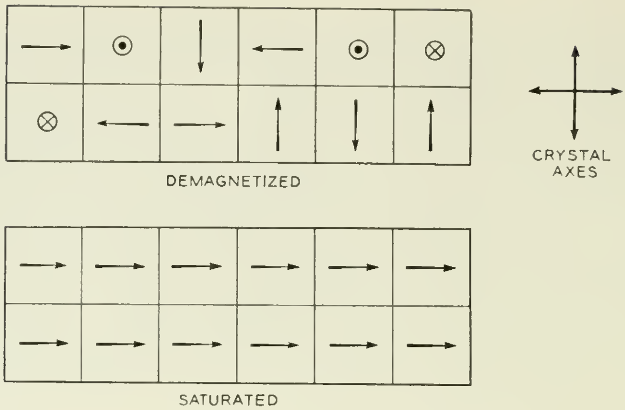
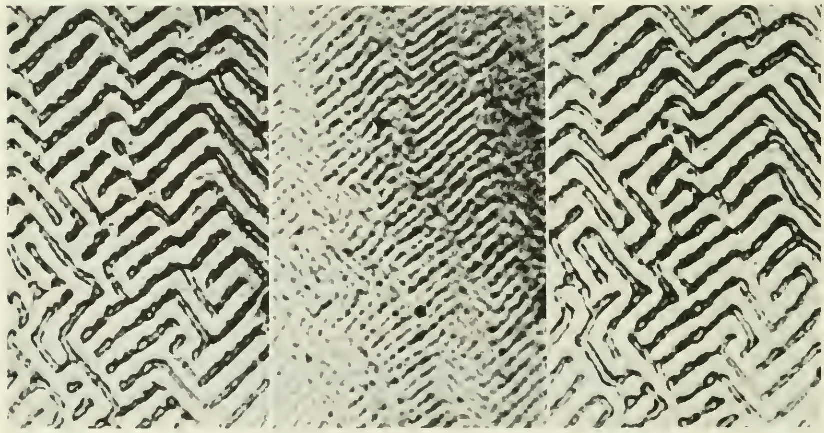


Fig. 8—Domains in a single crystal of iron.

increasing the resultant magnetization in the direction of the field. These changes take place suddenly, and they are the cause of the Barkhausen effect; each sudden change in orientation of a domain accounts for one step in the magnified magnetization curve shown in Fig. 5, or for one click heard in the telephone receiver when listening to the Barkhausen effect.

There is even more direct evidence of the existence of domains in a piece of iron. The iron is placed under a microscope with a magnification of 500, and is covered with a colloidal suspension of iron oxide.[29] It is found (Fig. 9) that the colloidal particles are concentrated along



(a) (b) (c)

Fig. 9—Powder-patterns for iron (McKeehan and Ellmore): (left) field outward; (middle) demagnetized; (right) field inward.

lines determined by the crystal axes, indicating that stray magnetic fields go in and out of the surface just as if some sections were magnetized differently from their neighbors. This occurs even when the iron is unmagnetized, but never occurs with materials that are not ferromagnetic.

Now consider in more detail by what processes changes in magnetization occur. Most changes are attributable to the reorientation of electron spins in domains, from one direction of easy magnetization to another (Fig. 10). These are the changes that take place over the

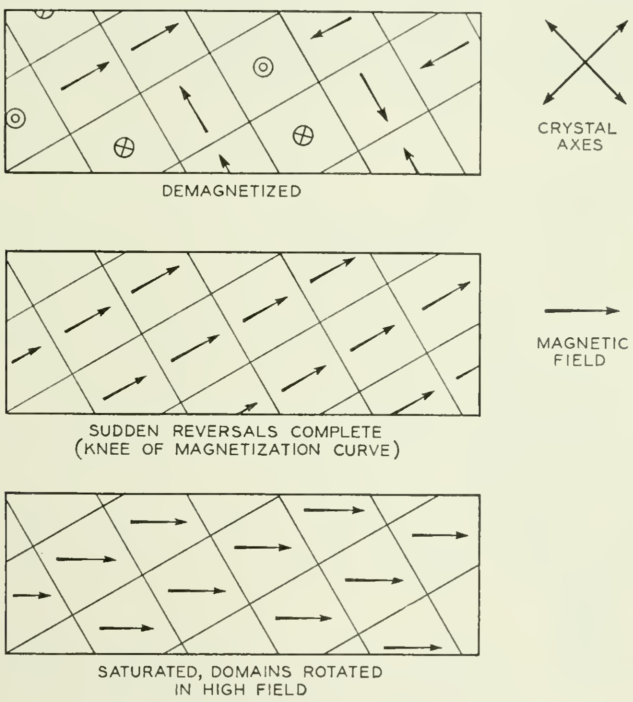


Fig. 10—As the magnetic field strength increases, domains first change direction suddenly, then rotate smoothly.

large central portion of the magnetization curve. In general, however, it is obvious that this process is complete before the material is saturated. When all the domains are magnetized parallel to that direction of easy magnetization which is nearest to the direction of the applied field, the only way in which the magnetization can be increased further is by rotating the electron spins in each domain out of the stable position toward the field direction. Such a process is described loosely as the "rotation of the domain." This is the process that

occurs in high fields, of the order of 10 to 100 œrsteds; as may be seen in Fig. 7, its beginning corresponds to the place where the curves suddenly bend over, away from the almost vertical section. It is only when the field is applied to a single crystal in the direction of easiest magnetization that this last process is avoided. When the field is applied in the direction of most difficult magnetization, the rotational process begins at a field-strength lower than in any other case.

One other important property of single crystals is accounted for by this picture. This property is evident when a field is applied to a single crystal in a direction not parallel to a principal axis. For example, let the field be applied 30 degrees from a cubic axis of an iron crystal, as indicated in Fig. 11 by the longest arrow. As this field is

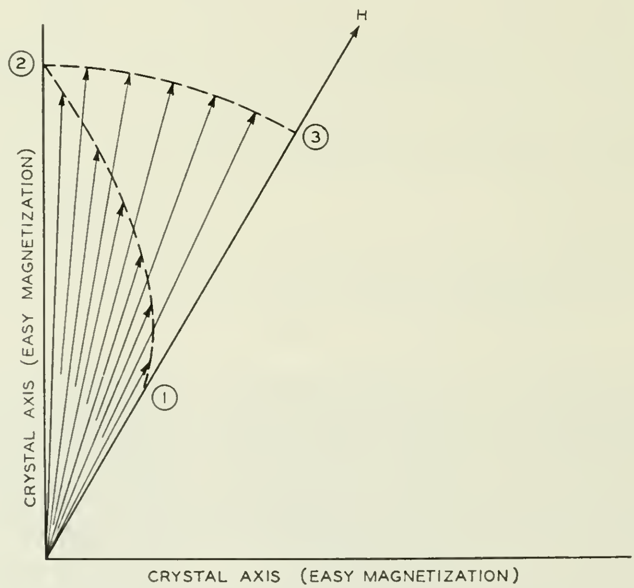


Fig. 11—Vectors represent *B-H* in iron, increasing in magnitude as the magnetic field (*H*) increases. First *B-H* is parallel to *H* (1); then as *B-H* increases it deviates in direction from *H* (2); and finally in high fields is again parallel to *H* (3).

increased from zero, the magnetization will correspond in magnitude and direction to the other arrows shown. First it is parallel to the field, but as the field increases it deviates toward a direction of easy magnetization until finally it is saturated in that direction. As the field is increased further, the magnetization approaches again the direction of the field and finally is saturated in this direction. Theory

agrees with experiment in that it predicts [30] the direction and amount of the deviation of B from H for any given value of B.

The two ways of changing magnetization that have been described for single crystals, namely sudden changes to new directions of easy magnetization, and continuous rotation of domains, apply equally well to ordinary polycrystalline material, the properties of the latter being those of the former averaged for all orientations. One result of this averaging, of course, is that the specimen is now isotropic and B is parallel to H.

These last remarks must be qualified, for the magnetic materials used by engineers are not always isotropic, that is, the crystal axes are not always distributed equally in all directions. It has been known for many years that when a metal sheet is rolled, the crystals composing it tend to be oriented in special ways with respect to the direction of rolling and to the rolling plane. Even after the sheet has been annealed and recrystallized, these special orientations exist, in some metals all the way up to the melting point. Since the magnetic properties depend on the crystal direction in a single crystal, it follows that sheets composed of crystals having special orientations will not have the same magnetic properties in all directions. This was observed some years ago in iron, nickel, and iron-nickel alloys.[31] More recently, there has appeared on the market a silicon-iron [32] alloy for which the permeabilities in different directions are markedly different. Measured parallel to the direction of rolling this material has a permeability in high fields (B = 15,000 gausses) of 4,000, while measured at right angles to the direction of rolling the permeability is only 400. X-ray analysis shows [33] that the crystals in this material are aligned so that most of them have a cubic axis lying within a few degrees of the direction of rolling. Thus the direction of rolling coincides with the direction of easy magnetization.

In considering the properties of single crystals, the properties in very low fields have not been considered, chiefly because precise data for single crystals are very difficult to obtain. The process that occurs in this region in single crystals and polycrystals must be different from either of the two so far considered, because in ordinary polycrystalline material there are no discontinuities in magnetization, i.e., no Barkhausen effect, and also the fields are not strong enough to rotate the domains to any significant extent against the crystal forces, out of a direction of easy magnetization. Knowing the relation between magnetic force and angular displacement in high fields, it is calculated that if this same mechanism applied to changes in magnetization in very low fields the highest value of initial permeability in iron would

be about 20 instead of many thousands. In the past the process occurring in low fields has been the cause of much speculation, but recently a satisfactory explanation seems to have been found.[5] The changes that take place here are visualized as displacements of the boundaries of domains (Fig. 12); the transition region of a few atom

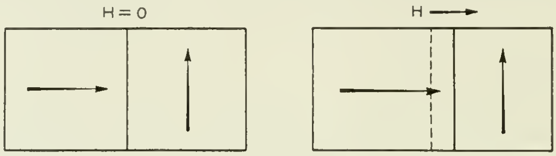


Fig. 12—Magnetization in very low fields progresses by slight displacement of domain boundaries (Becker).

diameters (calculated from the forces of exchange to be about 30 atom-diameters [7]) moves so as to enlarge a domain magnetized in the direction of the field at the expense of a domain pointing in a less favorable direction. Such a movement can progress for only a short distance compared with the linear dimensions of a domain, and is limited by the strains present in any actual material.

Thus in the magnetization of an ordinary well-annealed ferromagnetic material three processes occur, corresponding to the three well known sections of the magnetization curve (Fig. 13): growth of one

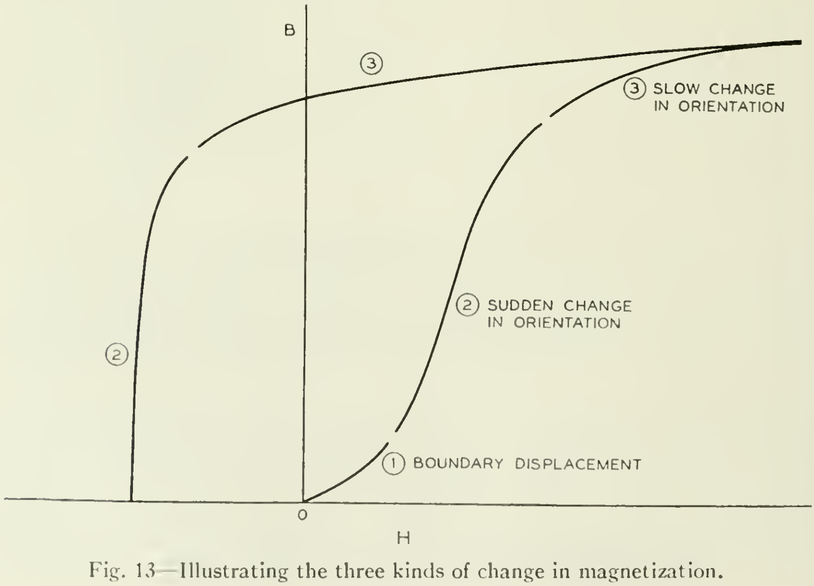


Fig. 13—Illustrating the three kinds of change in magnetization.

domain at the expense of a neighboring one in the initial portion of the curve, sudden changes of direction of domains (with resulting large energy losses) in the middle portion, and continuous or smooth rotation of the domains in the upper portion. The latter two processes occur during the traversal of a large hysteresis loop with tips at high flux densities; the first process is important only in low fields after demagnetization.

Effect of Strain

This picture of the changes in magnetization has been made for materials that are free from any considerable strain. As a matter of fact, strain can affect magnetization in an important way, and under certain circumstances a tensile stress of 5,000 pounds per square inch may change the flux density B as much as 10,000 gausses [34]—almost from zero magnetization to saturation (Fig. 14). The effect is il-

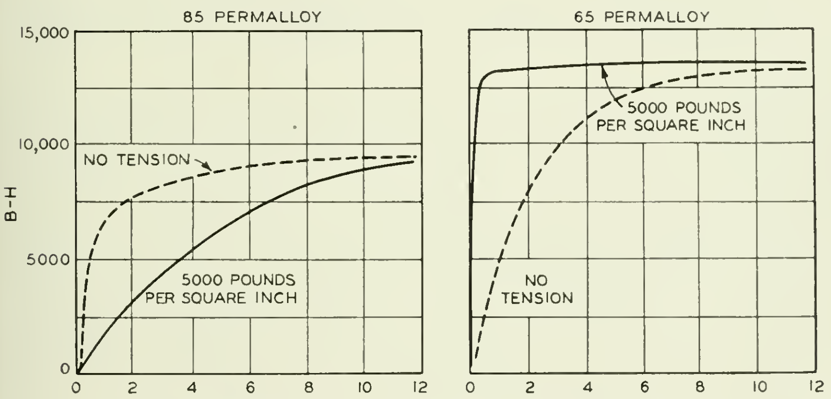


Fig. 14—Effect of tension on magnetization (Buckley and McKeehan).

lustrated well by data for 65 and 85 permalloy (iron-nickel alloys containing, respectively, 65 and 85 per cent nickel). For 65 permalloy the effect of tension is to increase the magnetization in all fields; for 85 permalloy the effect is the opposite; and in each case the effect of compression is opposite to that of tension. For ordinary iron the effect of tension is to increase the magnetization in small fields, but to decrease it in high fields.

The effect of strain on magnetization has its counterpart in an effect of magnetization on the length of a piece of ferromagnetic material. When a rod of iron is magnetized its length increases by a small amount. This is but one example of a large class of effects exhibited

by all ferromagnetic bodies and known collectively as "magnetostriction." Figure 15 shows the data for change in length of rods of nickel, iron, and two alloys, plotted against the field H on the one hand and against relative B-H on the other. When saturation of magnetization is reached, the limiting value of magnetostriction, called "saturation magnetostriction," also is attained. Its values for some

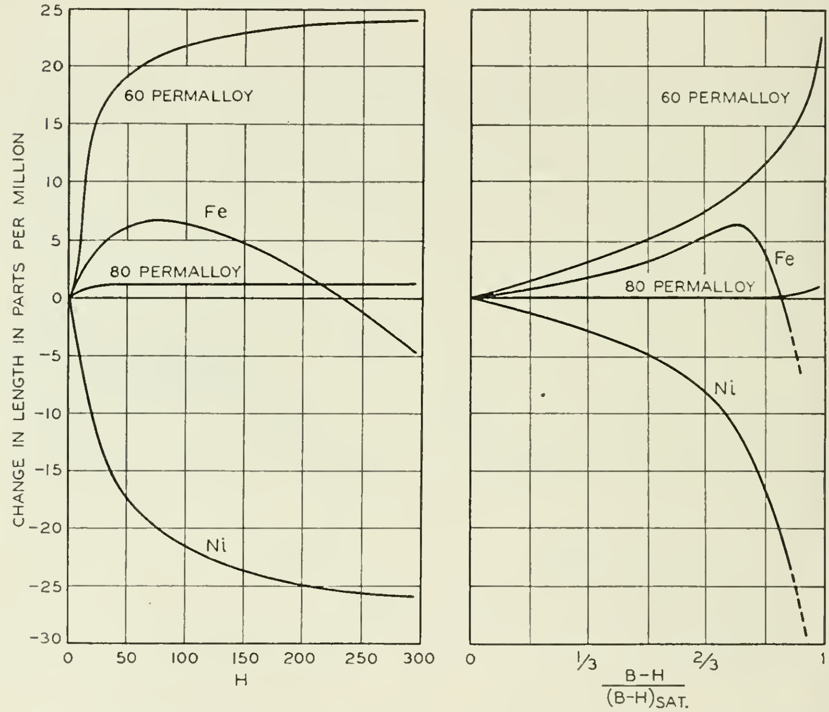


Fig. 15—Magnetostriction in iron, nickel, and two iron-nickel alloys (permalloys).

iron-nickel alloys are shown in Fig. 16. Note here that the change in length is an extension in the alloys containing less than 81 per cent nickel, a contraction otherwise. There is a close relation between magnetostriction and the effect of strain on magnetization, it being a general rule that when the magnetostriction is positive (increase in length with magnetization) the effect of tension is to increase magnetization, and vice versa (Figs. 14, 15, and 16).

How much can theory say of magnetostriction and the effect of strain on magnetic properties? Figure 17 shows how the atoms are arranged in an iron crystal; each atom here is supposed to have a

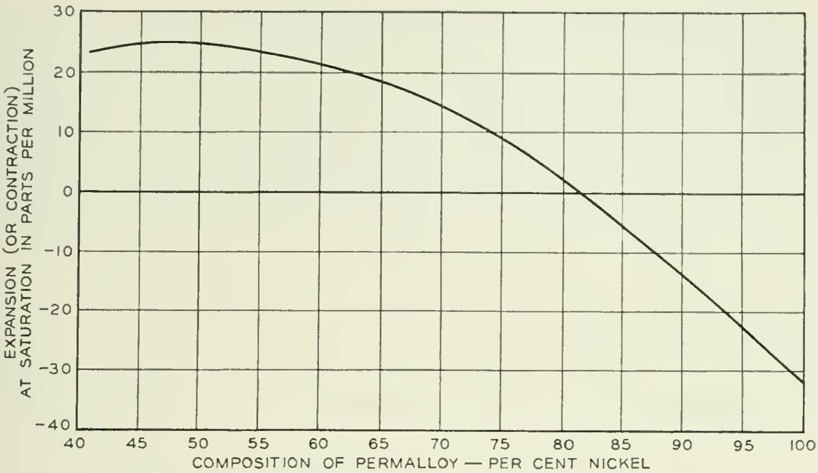


Fig. 16—Saturation magnetostriction of the permalloys (McKeehan and Cioffi, Schulze).

definite magnetic moment as a result of the spin and orbital motion of the electrons. This supposition makes it possible to calculate the magnitude of the mutual magnetic forces which are opposed by the elastic forces holding the crystal together. For iron, the calculations [5] indicate that equilibrium is reached when there has been a slight increase in length in the direction of magnetization and a decrease in length at right angles to this direction such that the volume remains practically unchanged. This calculated magnetostriction is in agreement with experiment as to sign and order of magnitude. With nickel the agreement is not so satisfactory. But in each case the theory is clear in predicting the proper qualitative relationship between magnetostriction and change in magnetization caused by strain.

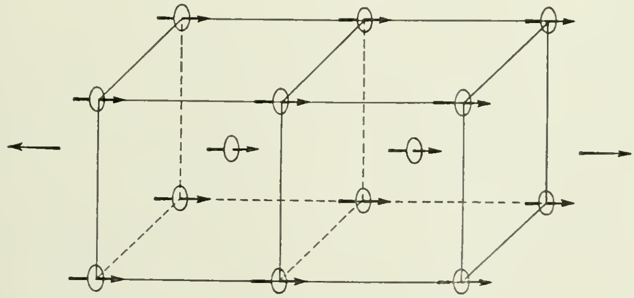


Fig. 17—The magnetic forces between atoms cause a slight elongation in iron (magnetostriction).

Thus magnetostriction and the magnetic effects of strain are reciprocal properties, and result from the same kind of magnetic forces between atoms as those that account for the variation in magnetic properties in different directions in a crystal. Just as the crystal structure determines a direction of easy magnetization in a strain-free crystal, so the strain controls the direction of easy magnetization when the strain is sufficiently great. Figure 18 shows how the domains are

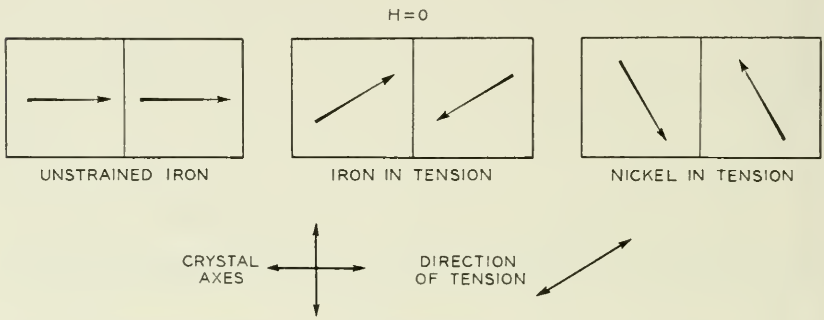


Fig. 18—Domains are oriented by crystal forces and by strain; $H = 0$.

magnetized parallel to the crystal axes in unstrained iron, and how a sufficiently large tension will orient the magnetization parallel to the direction of tension in iron and at right angles to the direction of tension in nickel. When the stress is as large as 10,000 to 30,000 pounds per square inch, the strain effect begins to predominate over the crystal effect and the direction of magnetization is determined mainly by the strain. The calculations show also that in a material having positive magnetostriction the magnetization is increased by tension. In a qualitative way these considerations explain the increase in permeability of 65 permalloy (having positive magnetostriction) and the decrease in 85 permalloy (with negative magnetostriction). But so far the theory is quite inadequate to predict the magnitude of the effect.

In addition to uniaxial homogeneous strains, such as those produced by stretching a wire in the direction of its length, random (heterogeneous) strains are often found that vary in magnitude, sign, and direction from point to point throughout a material. Such strains are produced by cold working, phase transformations, and the like. In such materials the direction of magnetization in a domain is determined by the local strain, and is more stable the larger the strain. So it can be appreciated that it is harder to change the magnetization of a material that is more severely hard worked. These internal strains are the same ones that contribute to the hardness of a metal—hence the parallelism between magnetic hardness and mechanical hardness, which is so well known.

This relation between internal strain and permeability is illustrated by the data [35] shown in Fig. 19. The permeabilities of a series of specimens of 70 permalloy tape, originally cold rolled, increase as the annealing temperature is raised. X-ray data (the angular width of the reflected X-ray beam) on these same specimens indicate the magnitude of the internal strains existing, and show that they become progressively less as the annealing temperature is increased, the most rapid change taking place in each case between 400 and 600 degrees centi-

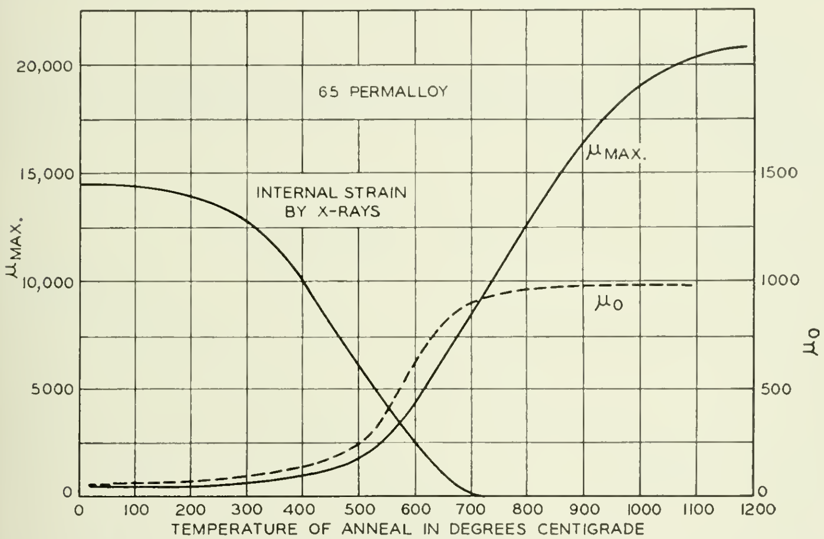


Fig. 19—Magnetic permeability rises as internal strain is diminished by annealing (Dillinger and Haworth).

grade, in which region the microscope shows recrystallization has occurred.

Following out this same idea, it may be surmised that to make good material for a permanent magnet something with very intense internal strains is required. The direct determination by X-rays of internal strain in a good permanent magnet, confirms this view (Fig. 20). Here the widths of the reflected X-ray beams directly measure the internal strains. For comparison with the permanent magnet material, curves are shown for other materials with less internal strain. The magnet material in this case was an iron-nickel-aluminum alloy that was precipitation-hardened, a method used more and more extensively during the last three or four years for such materials. This method is often applicable when the alloy [36] in the stable condition consists of two phases at room temperature (Fig. 21), but when at a

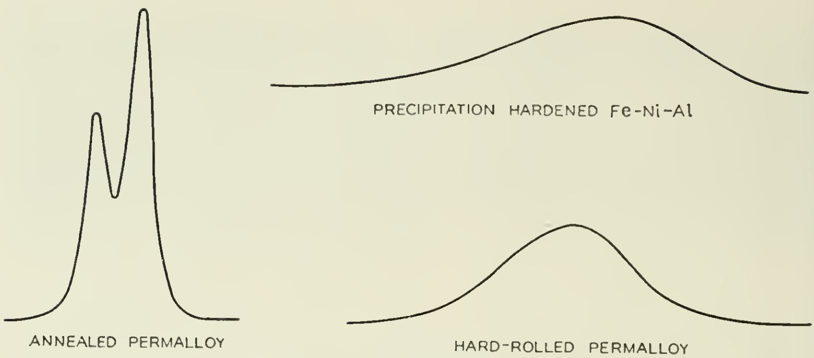


Fig. 20—The width of X-ray reflections indicates the amount of internal strain. Ordinates, intensity of X-rays reflected from metal surface; abscissas, angle of reflection.

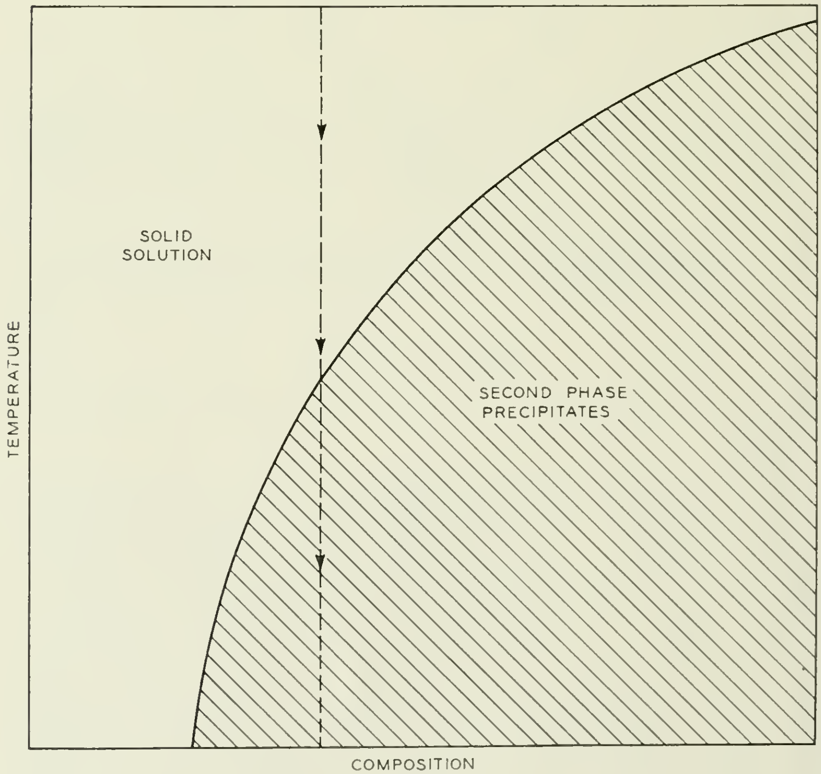


Fig. 21—Precipitation hardening of an alloy for a permanent magnet, such as an alloy of iron, nickel, and aluminum.

higher temperature the one phase dissolves completely in the other to form a solid solution. In making the material, it is quenched rapidly from a high temperature and then reheated to 700 degrees centigrade, at which point the second phase precipitates slowly in very finely divided form. When the optimum amount has precipitated, the material is cooled to room temperature, no more changes occurring. Each submicroscopic precipitated particle is a center of strain, and it is the presence of these unusually large internal strains that is responsible for the good quality of the permanent magnet.

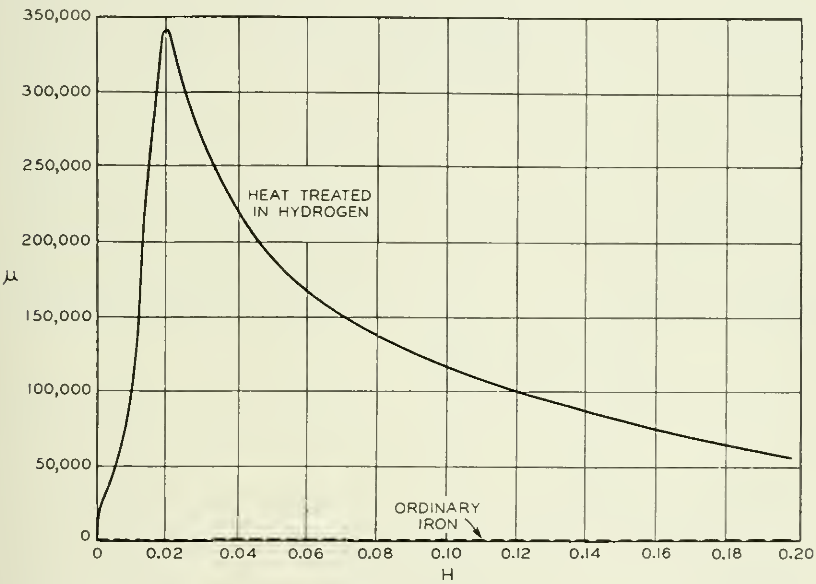


Fig. 22—Permeability curves of ordinary iron and of iron purified by heat treatment in hydrogen at 1,500 degrees centigrade (Cioffi).

Going now to the other extreme, where ease of magnetization is required, it is known, of course, that thorough annealing and a homogeneous structure are beneficial. Still there are at least two sorts of strains that annealing will not relieve. One is that attributable to the non-metallic chemical impurities that do not fit into the regular arrangement of atoms in a pure metal or alloy. It has been found recently that by heat treating iron in hydrogen at about 1,500 degrees centigrade the non-metallic impurities are largely removed, and that what are called "chemical strains" are much reduced. As a result it is found (Fig. 22) that the maximum permeability increases from 10,000 to 340,000,[12] and a large reduction in mechanical hardness occurs simultaneously.

After the chemical strains and the strains resulting from cold working have been removed, there is still another kind of residual strain—that attributable to magnetostriction. These are ordinarily random in direction because they are associated with randomly oriented domains, but by a suitable trick they all can be oriented so as to favor magnetization in a single desired direction at the expense of ease of magnetization at right angles. This trick is heat treatment in the presence of a magnetic field. Without going into a more detailed explanation, the experimental results obtained [15] about two years ago will be given.

When an annealed specimen of 65 permalloy is heated for a few minutes at 650 degrees centigrade while it is subjected to a magnetic field of 10 œrsteds, the maximum permeability is increased from about 20,000 to over 600,000 as shown in Fig. 23. This material holds the

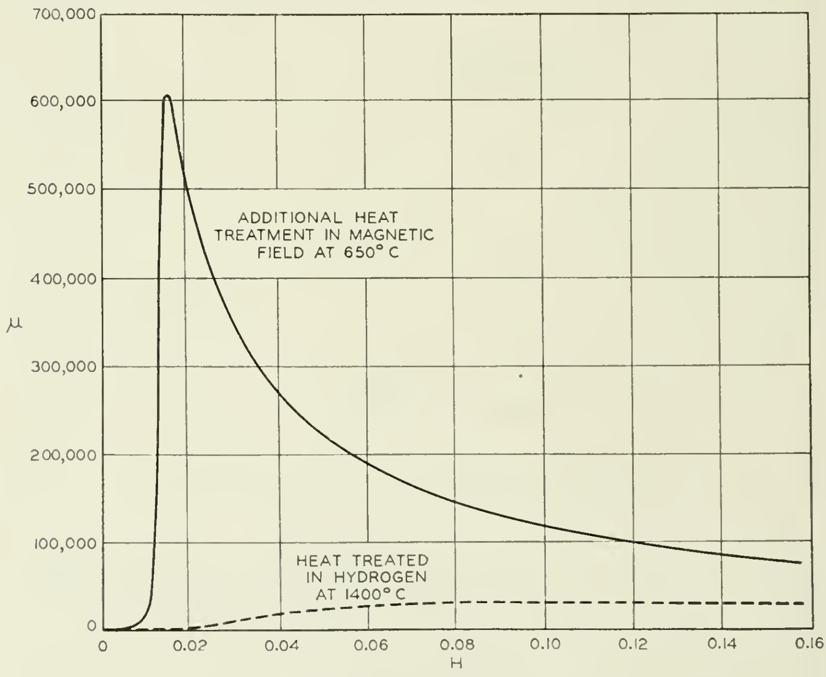


Fig. 23—Permeability curves of 65 permalloy after heat treatment in hydrogen and additional heat treatment in a magnetic field (Dillinger and Bozorth).

records for the highest maximum permeability, the lowest coercive force, and the lowest hysteresis loss at high flux densities. It may be compared with the most permeable material known in 1900, iron with a maximum permeability of less than 3,000.

So far only the effects of stress on the orientation of domains in medium and high fields have been considered. But stress has an effect on the initial permeability also. It has been said already that in very weak fields a change in magnetization is attributed to a movement of the boundaries between domains, the domains oriented nearly parallel to the field growing at the expense of adjacent domains oriented in less favorable directions. Such a growth obviously may be hindered by strain. A relation has been derived [37] connecting the initial permeability with the internal stress and other magnetic quantities:

$$\mu_0 = \frac{0.018(B - H)^2_{\text{sat.}}}{(\Delta l/l)_{\text{sat.}}\,\sigma_i},$$

where μ_0 is the initial permeability, $(B - H)_{\text{sat.}}$ and $(\Delta l/l)_{\text{sat.}}$ are the (ferric) induction and magnetostriction at saturation, and σ_i is the average value of the internal stress in dynes per square centimeter.

Even when there are no internal strains caused by impurities, insufficient annealing, etc., there generally will be the strains of magnetostriction itself, and these will hinder the growth of one domain at the expense of another (Fig. 24). In this case the stress in the foregoing

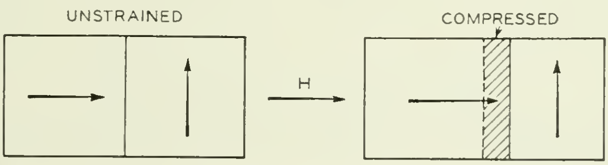


Fig. 24—Magnetostriction in the shaded region acts as a barrier to further change in magnetization.

equation is equal to Young's modulus, E, multiplied by the magnetostrictive strain,

$$\sigma_i = E(\Delta l/l)_{\text{sat.}}$$

and the former equation becomes

$$\mu_0 = \frac{0.018(B - H)^2_{\text{sat.}}}{(\Delta l/l)^2_{\text{sat.}}\,E},$$

This equation really gives a theoretical upper limit for μ_0. These theoretical limits and the highest observed values for iron-nickel alloys are shown in Fig. 25. This indicates why the composition of the "permalloy" having the highest initial permeability is very nearly the same as that for which the magnetostriction is zero.

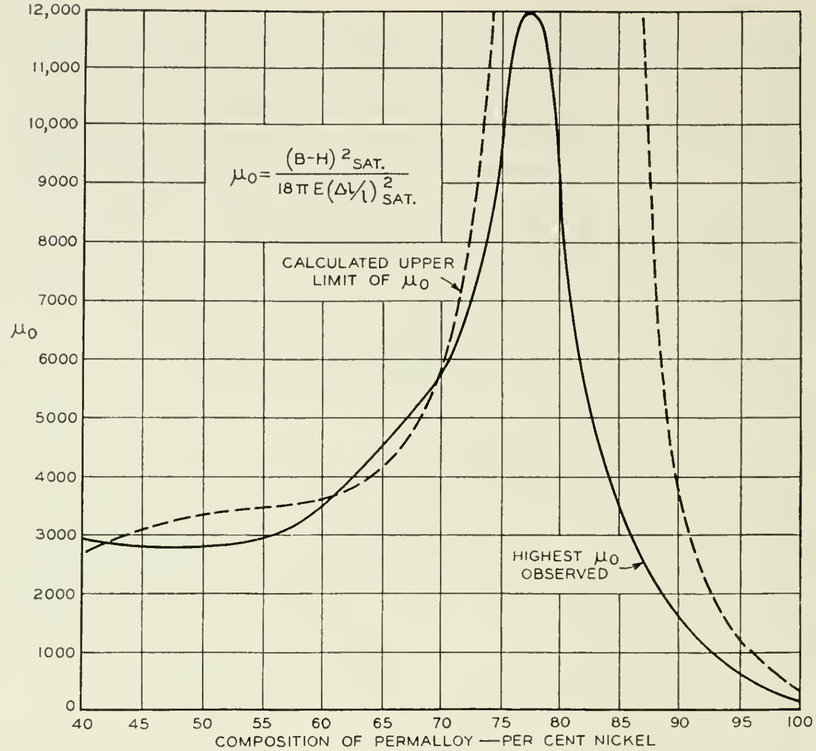


Fig. 25—Comparison of the theoretical upper limit of initial permeability (Kersten) with the highest initial permeabilities observed for iron-nickel alloys (Arnold and Elmen, Schulze).

The effects of strain will now be summarized briefly. The origin of the effects lies in the *magnetic* action between neighboring atoms. The magnetic action is balanced by the elastic (electrostatic) forces between atoms. The balance of these forces results in a change in shape of the magnetic body when it is magnetized (magnetostriction), and also a change in magnetization resulting from strain (strain-sensitivity). Magnetization may be either aided or hindered by a homogeneous uniaxial strain, the effect depending on the magnetostriction in a way that can be estimated qualitatively but not quantitatively. But material in which local strains are directed at random is more difficult to magnetize because the strains prevent a change in magnetization; and the more intense such strains are, the harder the material is to magnetize or demagnetize. The effect of local strains upon the initial permeability can be calculated with fair success, but other magnetic quantities, such as maximum permeability, can as yet be estimated in a qualitative way only.

Summary

In concluding the author wishes to go back from here to summarize what is known about the origin of the forces responsible for the various magnetic properties and about the sizes of the various units. This information is summarized in Table II.

TABLE II

Summary of Data Regarding Origin of Forces Responsible for Various Magnetic Properties

Unit Concerned	Property	Origin of Property	Size of Magnetic Unit
Electron	Magnetic moment	Electron spin	One unit of spin per electron
Paramagnetic atom	Magnetic moment	Uncompensated spins and orbital motions of electrons	4, 3, and 2 uncompensated spins per atom in Iron, Cobalt and Nickel, respectively
Domain	Ferromagnetism. Change in properties at Curie point	"Exchange" between electrons in neighboring atoms	Volume of domain is about $(0.001\ \text{inch})^3$
Single crystal or region of homogeneous strain	Crystal anisotropy. Magnetostriction. Strain sensitivity	Magnetic forces between atoms	10^8 domains per cubic centimeter
Polycrystal	Orientation-average of single crystals and strain units	Sum of effects of single crystals and strains	Size of specimen

Acknowledgment

I take pleasure in acknowledging the detailed criticisms given by Dr. K. K. Darrow, and also the more general criticism given by Dr. O. E. Buckley during the preparation of this paper.

References

Most of the material used in this article can be found in more technical form in the following books and summarizing articles. For some topics reference must be made to the original works, as indicated in the text.

Books

1. "Magnetism and Matter," E. C. Stoner. Methuen, London, 1934.
2. "Magnetic Induction in Iron and Other Metals," J. A. Ewing. Electrician, London, 3d ed., 1900.
3. "Magnetic Properties of Matter," K. Honda. Syokwabo, Tokyo, 1928.
4. "Die ferromagnetischen Legierungen," W. S. Messkin and A. Kussmann. Springer, Berlin, 1932.

Recent Summarizing Articles

5. "Theorie der Elektricität," R. Becker. Teubner, Leipzig, chap. 4, v. 2, 1933.
6. "Magnetisms," R. Becker and R. Landshoff, *Die Physik*, **3**, No. 2, 1935, p. 100–8.
7. "Handbuch der Radiologie," F. Bloch. Akad. Verlag., Leipzig, 2d ed., v. 6, pt. 2, 1934, p. 604.
8. "Handbuch der Physik" (Geiger u. Scheel), H. Bethe. Springer, Berlin, 2d ed., v. 24, pt. 2, 1933, p. 585.
9. "Stand der Forschung und Entwicklung auf dem Gebiet der ferromagnetischen Werkstoffe," A. Kussmann. *Archiv. für Elektrotechnik*, **29**, May 1935, p. 297–32.
10. "Magnetic Materials," W. C. Ellis and E. E. Schumacher. *Metals and Alloys*, **5**, Dec. 1934, p. 269–74, and **6**, Jan. 1935, p. 26–9.

Original Sources

11. "The Magnetic Properties of Some Iron Alloys Melted in Vacuo," T. D. Yensen. *A. I. E. E. Trans.*, **34**, 1915, p. 2601–41.
12. "New High Permeabilities in Hydrogen Treated Iron," P. P. Cioffi. *Phys. Rev.*, **45**, May 15, 1934, p. 742; and unpublished data.
13. These are without additions of other elements except for deoxidation. With additions of copper and molybdenum, material with initial permeability as high as 40,000 has been reported by H. Neumann. *Arch. f. tech. Mess.*, **4**, Dec. 1934, T-168.
14. "The Magnetic and Electrical Properties of the Iron-Nickel Alloys," C. F. Burgess and J. Aston. *Met. and Chem. Engg.*, **8**, Jan. 1910, p. 23–6.
15. "Heat Treatment of Magnetic Materials in a Magnetic Field," J. F. Dillinger and R. M. Bozorth. *Physics*, **6**, Sept. 1935, p. 279–91. See also earlier paper "Permeability Changes in Ferromagnetic Materials Heat Treated in Magnetic Fields," G. A. Kelsall. *Physics*, **5**, July 1934, p. 169–72.
16. G. Panebianco, Rend. d. Napoli, **16**, 1910, p. 216.
17. "Permalloy, an Alloy of Remarkable Magnetic Properties," H. D. Arnold and G. W. Elmen. *Franklin Inst. Jour.*, **195**, May 1923, p. 621–32.
18. "Magnetic Properties of Perminvar," G. W. Elmen. *Franklin Inst. Jour.*, **206**, Sept. 1928, p. 317–38.
19. "Saturation Value of the Intensity of Magnetization and the Theory of the Hysteresis Loop," E. H. Williams. *Phys. Rev.*, **6**, Nov. 1915, p. 404–9.
20. "The Iron Cobalt Alloy, Fe_2Co, and Its Magnetic Properties," T. D. Yensen. *Gen. Elec. Rev.*, **19**, Sept. 1915, p. 881–7.
21. "New K. S. Permanent Magnetic" (in English), K. Honda, H. Masumoto, and Y. Shirakawa. *Science Rep., Tohoku Imp. Univ.*, **23**, 1934, p. 365–73.
22. S. J. Barnett, *Rev. Mod. Physics*, **7**, 1934, p. 129.
23. "Atomic Shielding Constants," J. C. Slater. *Phys. Rev.*, **36**, July 1930, p. 57–64. See also ref. 8.
24. "Zur Theorie des Ferromagnetismus," W. Heisenberg. *Zeit. Physik*, **49**, July 1928, p. 619–36.
25. In a recent paper G. Urbain, P. Weiss and F. Trombe, *Compt. Rend.*, **200**, 1935, p. 2132, report that the rare-earth metal gadolinium is ferromagnetic. For this substance in the metallic state D/d is about 3, in accordance with Slater's rule (see Fig. 3). This usually high value of D/d indicates that gadolinium should have a low Curie point, which in fact is observed to be 16 degrees centigrade.
26. "Barkhausen Effect, II.—Determination of the Average Size of the Discontinuities in Magnetization," R. M. Bozorth and J. F. Dillinger. *Phys. Rev.*, **35**, April 1, 1930, p. 733–52.
27. "Zur Theorie der Magnetisierungskurve von Einkristrallen," N. S. Akulov. *Zeit. Physik*, **67**, 1931, p. 794. See also "A Contribution to the Theory of Ferromagnetic Crystals," G. Mahajani. *Phil. Trans. Roy. Soc.*, **228A**, 1929, p. 63–114, and reference 5.
28. Here the ordinates are B-H instead of the more usual B, since the former approaches a limiting "saturation" value. B-H sometimes is referred to as "ferric induction."

29. "Surface Magnetization in Ferromagnetic Crystals," L. W. McKeehan and W. C. Elmore. *Phys. Rev.*, **46**, Aug. 1, 1934, p. 226–8, and private communication. See also the earlier experiments of L. v. Hamos and P. A. Thiessen, "Über die Sichtbarmachung von Bezirken verschiedenen ferromagnetischen Zustandes fester Korper," *Zeit. Phys.*, **75**, 1932, p. 562, and "Experiments on the Nature of Ferromagnetism," F. Bitter. *Phys. Rev.*, **41**, Aug. 15, 1932, p. 507–15.
30. "The Theory of the Ferromagnetic Anisotropy of Single Crystals," R. M. Bozorth. *Phys. Rev.*, **42**, Dec. 15, 1932, p. 882–9.
31. "Anisotropie in magnetischen Werkstoffen," O. Dahl and F. Pfaffenberger. *Zeit. Phys.*, **71**, Aug. 1931, p. 93–105.
32. "Discussion on the Influence of Grain Size on Magnetic Properties," N. P. Goss. *Trans. A. S. Metals*, **22**, Dec. 1934, p. 1133–9.
33. "The Orientation of Crystals in Silicon Iron," R. M. Bozorth, *Trans. A. S. Metals*, **23**, Dec. 1935, p. 1107–11.
34. "Effect of Tension Upon Magnetization and Magnetic Hysteresis in Permalloy," O. E. Buckley and L. W. McKeehan. *Phys. Rev.*, **26**, Aug. 1925, p. 261–73.
35. J. F. Dillinger and F. E. Haworth, unpublished data.
36. "Dauermagnetwerkstoffe auf der Grundlage der Auscheidungshärtung," W. Köster. *Stahl u. Eisen*, **53**, Aug. 17, 1933, p. 849–56.
37. "Über den Einfluss des elastischen Spannungszustandes auf die Grösse der Anfangspermeabilität," M. Kersten. *Zeit. techn. Physik*, **12**, 1931, p. 665–9, based on the earlier work "Zur Theorie der Magnetisierungskurve," R. Becker. *Zeit. Phys.*, **62**, May 1930, p. 253–69.

Some Equivalence Theorems of Electromagnetics and Their Application to Radiation Problems

By S. A. SCHELKUNOFF

After a review of the general aspects of the classical electromagnetic theory several "equivalence" theorems are established and illustrated with a number of examples from the diffraction theory. Then follows a discussion of possible applications of these theorems to radiation problems. The latter part of the paper is devoted to the calculation of the power radiated from an open end of a coaxial pair.

THE usual methods of calculating the power radiated by an electric circuit depend upon a determination of the electromagnetic field from the electric current distribution in the circuit. The best known of these methods consists in integrating the Poynting vector over the surface of an infinite sphere surrounding the circuit. This method has been used exclusively until recent years; to facilitate its application, John R. Carson obtained a compact general formula for the radiated power.[1] Another method[2] consists in calculating the work done against the forces of the field in supporting a given current distribution in the circuit. Theoretically either of the two methods is sufficient for solving any radiation problem. Practically, aside from inherent difficulties involved in the calculation of the electric current distribution in the first place, the preliminary integration for determining the field components E and H may be rather complex. Thus in obtaining the power radiated by a semi-infinite pair of perfectly conducting coaxial cylinders this preliminary integration has to be extended over the infinite surfaces of the two conductors. And yet by the Maxwell-Poynting theory, no energy can flow through the walls of the outer cylinders since the electric intensity E and hence the Poynting vector vanish there. Any energy which is radiated away must pass through the open end and it is natural to expect that there must be a method for calculating this energy from the conditions at the open end. The integration involved in this method would extend only over a comparatively small area of the open end. It is in search of a method of this type for calculating the radiated power that I was led some time ago to certain "equivalence theorems." Subsequently I learned that

[1] John R. Carson, "Electromagnetic Theory and the Foundations of Electric Circuit Theory," *The Bell System Technical Journal*, pp. 1–17, January 1927.

[2] A. A. Pistolkors, "The Radiation Resistance of Beam Antennas," *Proc. I. R. E.*, Vol. 17, No. 3 (1929). R. E. Bechmann, "On the Calculation of Radiation Resistance of Antennas and Antenna Combinations," *Proc. I. R. E.*, Vol. 19, p. 1471 (1931).

one of these theorems was discovered long ago, first by A. E. H. Love [3] and then by H. M. MacDonald [4] and proved by the latter [5] for the case of non-dissipative media in 1911. Another proof of this theorem, believed to be helpful from the physical point of view and extended so as to include the dissipative media, is given in this paper. After a brief review of some fundamental concepts we shall prove these equivalence theorems, discuss their significance, and solve one or two simple examples for illustrative purposes.

The physical sources of electromagnetic fields are electric and magnetic charges in motion, that is electric and magnetic currents. The radio engineer has never been interested in shaking magnets for the purpose of radiating energy and has settled into a habit of ignoring magnetic currents altogether as if they were non-existent. It is true that there are no magnetic conductors and no magnetic conduction currents in the same sense as there are electric conductors and electric conduction currents but magnetic convection currents are just as real as electric convection currents, although the former exist only in doublets of oppositely directed currents since magnetic charges themselves are observable only in doublets. And, of course, the magnetic displacement current, defined as the time-rate of change of the magnetic flux, is exactly on the same footing as the electric displacement current defined by Maxwell as the time-rate of change of the electric displacement. We shall find it convenient, at least for analytical purposes, to employ the concept of magnetic current on a par with the concept of electric current.

The two fundamental electromagnetic laws can now be stated in a symmetric form. Ampère's law as amended by Maxwell is: *An electric current is surrounded by a magnetic field of force; the induced magnetomotive force in a closed curve is equal to the electric current passing through any surface bounded by the curve.* In its original form, the "electric current" meant only the conduction current so that the law was applicable only to closed conduction currents. Maxwell's amendment consisted in including the displacement currents, thereby making the law applicable to open conduction currents. The second law is due to Faraday: *A magnetic current is surrounded by an electric field of force; the induced electromotive force in a closed curve is equal to the negative of the magnetic current passing through any surface bounded by the curve.* The rule for algebraic signs is as follows: choose some direction of the closed curve as positive and have an observer placed in such a way that

[3] A. E. H. Love, "The Integration of the Equations of Propagation of Electric Waves," *Phil. Trans.* A, Vol. 197, pp. 1–45 (1901).

[4] H. M. MacDonald, "Electric Waves," p. 16 (1902).

[5] H. M. MacDonald, "The Integration of the Equations of Propagation of Electric Waves," *Proc. London Mathematical Society*, Series E, Vol. 10, pp. 91–95 (1911).

this direction appears to him counterclockwise; then the positive direction of either the electric or the magnetic current is chosen *toward* the observer. If the currents are flowing toward the reader, the directions of the E.M.F. and the M.M.F. are as indicated in Fig. 1.

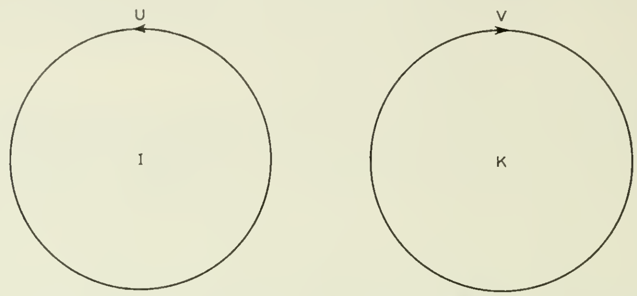


Fig. 1—The relative directions of the E.M.F. and the M.M.F. induced respectively by the electric current I and the magnetic current K are indicated by the arrows. Both I and K are directed toward the reader.

In the well-known way these two physical laws lead to a pair of partial differential equations

$$\operatorname{curl} E = -M, \qquad \operatorname{curl} H = J, \tag{1}$$

where J and M are respectively the total electric current density and the total magnetic current density. The electric density is composed of several parts; namely: the conduction current density, the displacement current density and the applied current density. The first of these components is, in many substances, proportional to the electric intensity E; the second is proportional to $\partial E/\partial t$; and the third is due to forces other than those of the field, mechanical or chemical, for instance. Similarly the magnetic current density is the sum of the magnetic displacement density proportional to $\partial H/\partial t$ and the impressed magnetic current density. Thus, we write

$$\operatorname{curl} E = -M_0 - \mu\frac{\partial H}{\partial t}, \qquad \operatorname{curl} H = J_0 + gE + \epsilon\frac{\partial E}{\partial t}, \tag{2}$$

where J_0 and M_0 are the densities of the impressed currents and the constants of proportionality g, ϵ and μ are respectively the conductivity, the dielectric constant and the permeability.[6]

The functions J_0 and M_0 are supposed to be known functions of coordinates and of time, representing the distribution of the physical

[6] A consistent practical system of units is used in this paper. Thus the E.M.F. is measured in volts, the electric current in amperes, E in volts per centimeter, H in amperes per centimeter, etc. The permeability of vacuum is then $4\pi 10^{-9}$ henries per centimeter and the dielectric constant $(1/36\pi)10^{-11}$ farads per centimeter.

sources in the space-time. If they are zero everywhere and at all times, the only physically significant solution of (2) must be $E = H = 0$ throughout the entire space and at all times.[7] If there are other solutions of (2), they are extraneous and some rule must be found for excluding them. Such extraneous solutions often find their way into mathematical equations because it is usually impossible to express *all* physical conditions by an equation or a system of equations. Naturally these remarks do not apply to a limited region of space or a finite interval of time. In fact, in many physical problems these "extraneous in the large" solutions of (2) can be advantageously used for expressing the general character of electromagnetic phenomena in a limited region and then obtaining, with the aid of the boundary and the initial conditions, the complete answer. But the philosophy of causality demands the dictum "no sources, no field" when considering the whole space-time. It may seem unnecessary to dwell at length on such obvious matters but they happen to be essential in the subsequent discussion if the arguments are to be taken as positive proofs rather than as plausible justifications.

Equations (2) are linear and the principle of superposition is applicable. This is in accordance with physical intuition which tells us that we can subdivide the impressed currents into elementary cells of volume dv, calculate the field due to a typical element, and obtain the total field by integration. For the typical element (2) becomes

$$\operatorname{curl} E = -\mu\frac{\partial H}{\partial t}, \qquad \operatorname{curl} H = gE + \epsilon\frac{\partial E}{\partial t}, \tag{3}$$

everywhere except in the infinitely small volume occupied by the element. The product of the current density and the volume of the element is called the *moment* of the element.

At times the impressed currents are confined to sheets so thin that their thickness can be disregarded without introducing a serious error in the result. This leads to a hypothetical infinitely thin *current sheet*. We pass from a real current sheet to an ideal one by assuming that the thickness of the former decreases and the current density increases in such a way that their product remains constant. This product is called the linear density of the sheet and it represents the current per unit length perpendicular to the lines of flow. The moment of a current element is now the product of the density of the sheet and the area of the element. Finally if the impressed current is confined to a

[7] We assume that all the electric and magnetic charges were originally in the neutral state, in which case their separation could be effected only through their motion. The argument could be extended so as to include purely static fields that may constitute an integral part of the universe but it is of no particular interest to us.

very thin filament, the moment is the product of the current and the length of the element.

It is the moment of the current element that determines its electromagnetic field. If the medium is non-dissipative, the actual expressions for the field components are obtained in terms of an auxiliary function called by Lorentz the *retarded magnetic vector potential.* For an electric current element of moment $p(t)$ this vector potential at any point P is parallel to the current density and is a function of the distance r from the element to P

$$A = \frac{p\left(t - \frac{r}{c}\right)}{4\pi r}. \tag{4}$$

The quantity c has the dimensions of a velocity and it appears that the action of the source travels outward with this velocity. But there is another solution of (3)

$$A_1 = \frac{p\left(t + \frac{r}{c}\right)}{4\pi r}. \tag{5}$$

One might wonder if this solution appertains in any way to the source; that is not the case, however. If the moment $p(t)$ is identically zero prior to some instant $t = t_0$, the field which can legitimately be attributed to the action of this source is also identically zero for any instant $t < t_0$. But (5) implies a non-vanishing field at distant points; it is as if the effect appeared before the cause. Any other solution is a combination of (4) and (5) and has to be rejected on the same grounds.

In terms of this auxiliary vector potential the field components can be expressed as follows

$$H = \operatorname{curl} A, \qquad \frac{\partial E}{\partial t} = \frac{1}{\epsilon} \operatorname{curl} H, \qquad E = \frac{1}{\epsilon} \int_{-\infty}^{t} \operatorname{curl} H \, dt. \tag{6}$$

If the moment is harmonic of frequency f, we regard it as the real part of $pe^{i\omega t}$. Then the vector potential and the field components are the real parts of the following expressions

$$A = \frac{pe^{-i\beta r}}{4\pi r}, \qquad H = \operatorname{curl} A, \qquad E = -i\omega\mu A + \frac{\operatorname{grad} \operatorname{div} A}{i\omega\epsilon}, \tag{7}$$

where the *phase constant*

$$\beta = \frac{\omega}{c} = \frac{2\pi f}{c} = \frac{2\pi}{\lambda},$$

λ is the wave-length, and the time factor $e^{i\omega t}$ is implied.

If the medium is dissipative, we have

$$A = \frac{pe^{-\sigma r}}{4\pi r}, \qquad H = \text{curl}\, A, \qquad E = -i\omega\mu A + \frac{\text{grad div}\, A}{g + i\omega\epsilon}. \tag{8}$$

The quantity σ is the *intrinsic propagation constant* of the medium and is defined by

$$\sigma = \sqrt{i\omega\mu(g + i\omega\epsilon)}. \tag{9}$$

In this case the action of the source at some point is not only delayed by the time needed for the disturbance to travel the intervening distance but also exponentially attenuated.

If instead of an electric current element, we are dealing with a magnetic element, the field components can be expressed in terms of an *auxiliary electric vector potential.* This vector F is given by

$$F = \frac{Pe^{-\sigma r}}{4\pi r}, \tag{10}$$

where the moment P of the element is the product of the magnetic current density and the volume of the element. The field components are then given by

$$E = -\text{curl}\, F, \qquad H = -(g + i\omega\epsilon)F + \frac{\text{grad div}\, F}{i\omega\mu}. \tag{11}$$

In the periodic case the general mathematical solution for the vector potential of an element is found to be a linear combination of any two of the following functions

$$\frac{e^{-\sigma r}}{r}, \qquad \frac{e^{\sigma r}}{r}, \qquad \frac{\cosh \sigma r}{r}, \qquad \frac{\sinh \sigma r}{r}. \tag{12}$$

All of these except the first become exponentially infinite at an infinite distance from the source and cannot be taken to represent the vector potential of a physical source. The last function is finite in any finite region; conceivably it can represent an electromagnetic field in the finite region free from physical sources. If the medium is non-dissipative it is impossible to exclude any of the solutions given by (12) on the grounds of their behavior at infinity—they all vanish there. But we may regard the non-dissipative case as the limit of the dissipative one and in this way establish a rule for finding the proper unique solution.

In the presence of a current sheet, equations (2) are valid on either side of it but not on it. Let us consider a cross-section of an electric current sheet, perpendicular to the lines of flow, and a curvilinear

rectangle $A'B'B''A''$ with two of its sides parallel to the sheet (Fig. 2). We assume that the current flows toward the reader and that $A'A''$ and $B'B''$ are vanishingly small. Since the M.M.F. around this rectangle is equal to the electric current passing through it and since this M.M.F. is merely the difference between the M.M.F.'s along the sides $A'B'$ and $A''B''$, we obtain

$$H_t' - H_t'' = J_l \tag{13}$$

by simply calculating these quantities per unit length of the rectangle. The tangential components of the magnetic intensity are regarded as

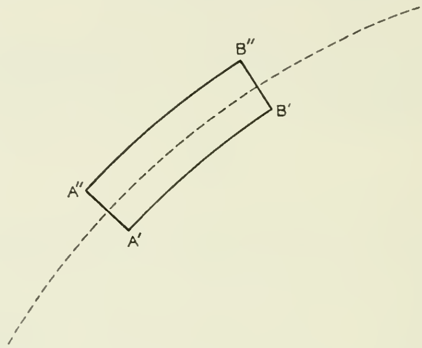


Fig. 2—A cross-section of a current sheet perpendicular to the lines of flow. The positive direction of the current is toward the reader.

positive when directed from A to B. Thus the tangential component of the magnetic intensity is discontinuous across an electric current sheet and the amount of the discontinuity is equal to the density of the sheet.

Similarly across a magnetic current sheet the tangential component of the electric intensity is discontinuous and the amount of this discontinuity is equal to the negative of the magnetic current density of the sheet; thus

$$E_t' - E_t'' = -M_l. \tag{14}$$

In deriving equations (2) it is also necessary to assume that g, μ and ϵ are continuous throughout the region under consideration. They have no meaning on the boundary between two different media. Since the boundary is a geometric surface, it cannot constitute either an electric or a magnetic current sheet. Hence the components of E and H tangential to such a boundary are continuous across it. These *boundary conditions* provide a link between the fields in the two media.

Let us suppose that we have a continuous distribution of sources on a closed surface C (Fig. 3) and that there are no other sources. We assume that the sources are harmonic of frequency f. The electromagnetic field $\mathfrak{F}$ produced by these sources can be calculated directly from this distribution with the aid of the above mentioned vector potentials. On the other hand, we can reason as follows.

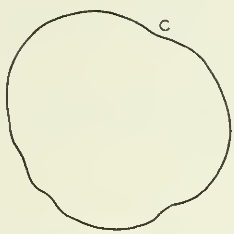


Fig. 3—A cross-section of a closed surface C.

There are no sources either inside or outside of C; hence everywhere except on C, we have

$$\text{curl } E = -i\omega\mu H, \qquad \text{curl } H = (g + i\omega\epsilon)E. \tag{15}$$

In the region inside of C we take that solution of (15) which is finite throughout this region and outside of C we select the solution vanishing at infinity. Both solutions will contain constants which can be determined from conditions (13) and (14) across the surface. The field $\mathfrak{F}'$ obtained in this manner is identical with $\mathfrak{F}$ because the difference $\mathfrak{F} - \mathfrak{F}'$ is everywhere source-free and thus must vanish.

Let us now reverse the process and, instead of starting with the known distribution of sources on C, suppose that we know the field and wish to find its sources. Let the known field $\mathfrak{F}$ be source free everywhere except on C. In order to determine these sources S we merely calculate the discontinuities in the tangential components of E and H across C. We can utilize this result to establish the major *Equivalence Principle.* For the outside portion of $\mathfrak{F}$ we can choose the outside portion of the field $\mathfrak{F}'$ produced by a given system of sources S' situated inside C and for the inside part of $\mathfrak{F}$ we take any field which is source-free there. The latter may be, for instance, the inside portion of the field $\mathfrak{F}''$ produced by some sources S'' situated outside C. Thus we arrive at the following *Equivalence Principle* discovered by Love and Macdonald [8]: a distribution of electric and magnetic currents on a given surface C can be found such that *outside* C it produces the

[8] See references 3 and 4 and also H. M. Macdonald, "Electromagnetism" (1934).

same field as that produced by given sources *inside* C; and also the field *inside* C is the same as that produced by given sources *outside* C. One of these systems of sources can be identically equal to zero.

The actual calculations are made as follows. From the discontinuities in the tangential components of E and H, we obtain J and M by (13) and (14). From these currents we find the two vector potentials

$$\begin{aligned} A &= \frac{1}{4\pi} \iint\limits_{(C)} \frac{J(x', y', z')}{r} e^{-i\beta r}\, dS, \\ F &= \frac{1}{4\pi} \iint\limits_{(C)} \frac{M(x', y', z')}{r} e^{-i\beta r}\, dS, \end{aligned} \tag{16}$$

where $r = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}$ is the distance between a point $P(x, y, z)$ somewhere in space and a point $P'(x', y', z')$ on C. From these potentials we calculate the electric intensity and the magnetic intensity by

$$\begin{aligned} E &= -\, i\omega\mu\, A + \frac{\text{grad div } A}{i\omega\epsilon} - \text{curl } F, \\ H &= \text{curl } A + \frac{\text{grad div } F}{i\omega\mu} - i\omega\epsilon\, F. \end{aligned} \tag{17}$$

The proof of the Equivalence Principle can be modified so as to throw some additional light on it. Let us suppose that given sources S' are inside the closed surface C and let us make our new synthetic field by obliterating the old field outside C and leaving everything as it was inside C. The new field has the same sources S' and besides it is discontinuous across C. These discontinuities are the additional sources S whose densities are calculable from (13) and (14). Since the new field is identically zero outside C, the field produced by S is such as to cancel the field produced by S' outside C. Thus the system of sources S acts as a perfect absorber for the electromagnetic wave produced by S'. Reversing the directions of the current distributions on C, we conclude that the system of sources $-S$ produces outside C exactly the same field as S'.

The Equivalence Principle is closely related to another theorem which we may call the Induction Theorem. Let us suppose that a closed surface C subdivides the entire space into two homogeneous media and that a system of sources S is given in one of those regions (Fig. 4). Let E, H be the field due to these sources on the assumption that the medium inside C is the same as that outside. The true field

outside C must vanish at infinity but it need not be the same as E, H; let it be $E + E'$, $H + H'$. The field E', H' must be source-free outside C. Inside C the field must be source-free; we shall designate it by E'', H''. The field E', H' is called the *reflected field* and E'', H'' the *refracted field.* The boundary conditions are such that the components of the electric and the magnetic intensities tangential to C must be continuous. Thus over the surface C, we have

$$\bar{E}_t + \bar{E}_t' = \bar{E}_t'', \qquad \bar{H}_t + \bar{H}_t' = \bar{H}_t''. \tag{18}$$

The bar over the letters is used to designate the values of the corresponding quantities on C. From (18) we obtain

$$\bar{E}_t'' - \bar{E}_t' = \bar{E}_t, \qquad \bar{H}_t'' - \bar{H}_t' = \bar{H}_t. \tag{19}$$

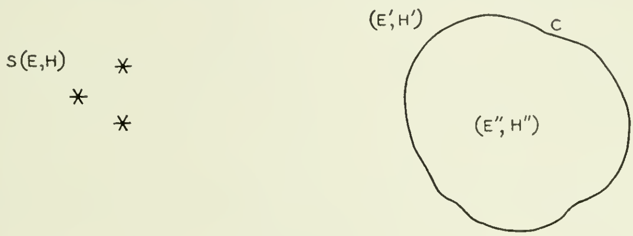


Fig. 4—The closed surface C is the boundary between two homogeneous regions in space. (E, H) designates the field produced by some system of sources S; (E', H') is the field reflected by the body C; and (E'', H'') is the field in the body.

Hence *the reflected and the refracted fields together constitute an electromagnetic field in the entire space; this field is source-free everywhere except on C and the distribution of sources on C is calculable from the given sources S.* This Induction Theorem is a generalization of the well-known theorem used in calculating the response of an electric circuit to an impressed field. Since the wires constituting the circuit are very thin, only the tangential components of E in the direction of the wires need be considered.

It may be noted that if the medium inside C is identical with that outside C, the "reflected field" must be absent and the "refracted field" must be identical with the field E, H due to the sources S. Thus the Induction Theorem leads to the Equivalence Principle.

The Equivalence Principle is evidently an extension of Kirchhoff's theorem. The latter deals with a single wave function instead of two vectors. Kirchhoff derived a formula for computing the wave function in the source-free region from its values and the values of its normal derivative over a closed surface separating the source-free region from

the region containing the sources of the wave functions. In the Theory of Sound the wave function represents the excess pressure or the velocity potential and Kirchhoff's theorem is valuable in the analysis of diffraction phenomena. Kirchhoff's theorem is also used in dealing with optical diffraction. We may also remark that Kirchhoff's formula is a mathematical expression of a principle governing compressional wave motion. This principle was first formulated by Huygens in the following form: each particle in any wave front acts as a new source of disturbance, sending out secondary waves, and these secondary waves combine to form the new wave front.[9]

Let us now examine one of the familiar diffraction problems in the light of the Equivalence Principle. Consider a source S and a perfectly absorbing screen (Fig. 5*a*). Such a screen will be defined in the usual manner: the impressed wave enters it without reflection but does not pass through it. If the screen is infinitely thin, this definition implies the existence of electric and magnetic currents in the screen whose densities are given by the postulated discontinuity in the field. In reality the "black bodies" absorb not by virtue of the coexistence of electric and magnetic currents but by virtue of electric currents alone with the aid of reflections taking place between atomic layers. The true mechanism of absorption is complex and requires more than a mere surface. In diffraction studies it has become a habit with us to ignore the precise nature of absorption and confine outselves to its implications; but it is just as well to know the nature of the ideal mechanism which we are substituting for the true mechanism.

We can apply the Equivalence Principle to the present problem in two ways. We can choose as our surface C a surface (1234) just on the other side of the screen. The part (23) contributes nothing; the equivalent distribution of sources S' over the parts (12) and (34) gives us a complete field to the right of the screen. On the other hand if S'' is the field due to the electric and magnetic currents in the screen induced by S, the total field is $S + S''$. The choice of the "surface C" that would yield this result is shown in Fig. 5*b* although the conclusion is obvious without recourse to the Equivalence Principle. Since to the right of C in Fig. 5*a* the two alternative fields must be the same, we have

$$S' = S + S'' \qquad \text{and} \qquad S' - S'' = S. \tag{20}$$

Incidentally the last equation is the expression for the Equivalence Principle as applied to S in the absence of the screen since $-S''$ is

[9] A. E. Caswell, "An Outline of Physics," p. 544 (1929).

the contribution of that portion of the equivalent layer which was removed by the screen.

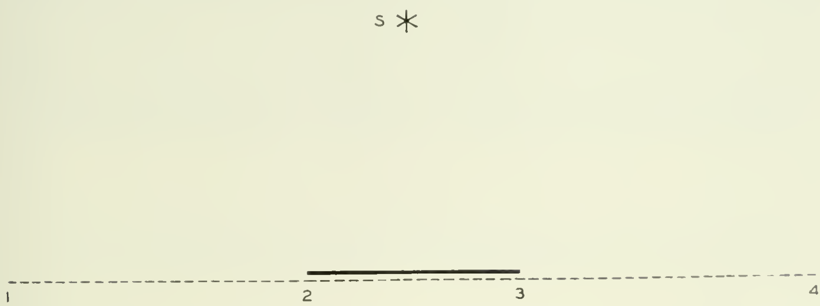


Fig. 5*a*—A source S in front of a screen the cross-section of which is shown in heavy lines.

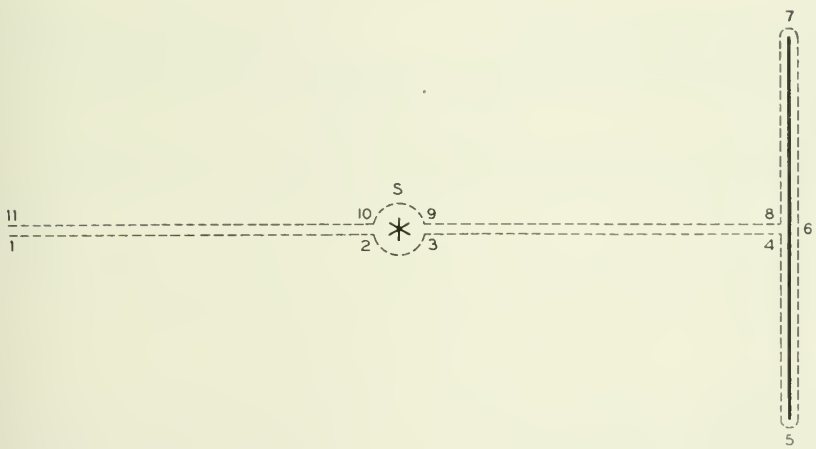


Fig. 5*b*—A source S in front of a screen the cross-section of which is shown in heavy lines.

The case of a hole in a perfectly absorbing screen (Fig. 6*a* and Fig. 6*b*) can be treated in the same manner and the reciprocity existing between this and the preceding case is quite evident. In terms of the sources previously defined the field to the right of the screen is $-S''$; by (20) this is the same as $S - S'$.

If the screen is a perfect conductor, the problem is much more complex. The screen will support electric currents but not magnetic currents. The densities of the electric currents are not calculable directly from the field S but from the condition that the component of the electric intensity tangential to the screen vanishes. The problem

is very difficult and its solution has been found in only a few special cases. It is true that once we know the electric currents in the screen, we can determine the field on both sides of the screen; but there is no simple way of calculating these currents exactly. Frequently it is assumed that, in so far as the side opposite to the source is concerned, a perfectly conducting screen is equivalent to a perfectly absorbing screen of the same geometric character. This is equivalent to a

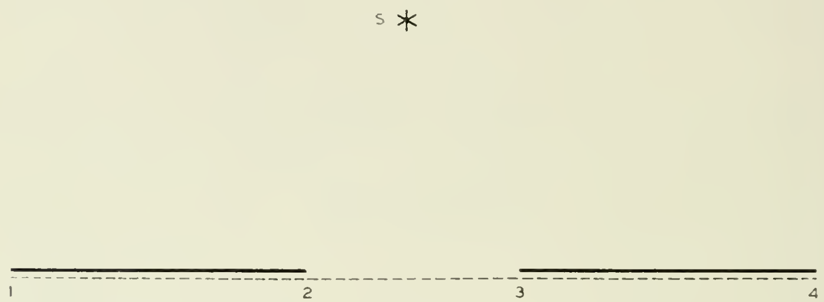


Fig. 6*a*—A source S in front of a screen the cross-section of which is shown in heavy lines.

Fig. 6*b*—A source S in front of a screen the cross-section of which is shown in heavy lines.

hypothesis that the electric current density of the screen is determined by the magnetic intensity impressed directly by the source S. We could take the results obtained from this hypothesis as a first approximation to the true results. The tangential component of the electric intensity calculated on the basis of this hypothesis does not vanish on the screen which means that we have violated the original hypothesis that the screen is a perfect conductor. If the discrepancy is not too great we might look for an additional electric current distribution to reduce this discrepancy.

There are times, however, when the current distribution in the "screen" can be determined with a fair accuracy without elaborate mathematics. It is so, for instance, in the case of a pair of perfectly conducting coaxial cylinders (Fig. 7*a* and Fig. 7*b*) in which the radii

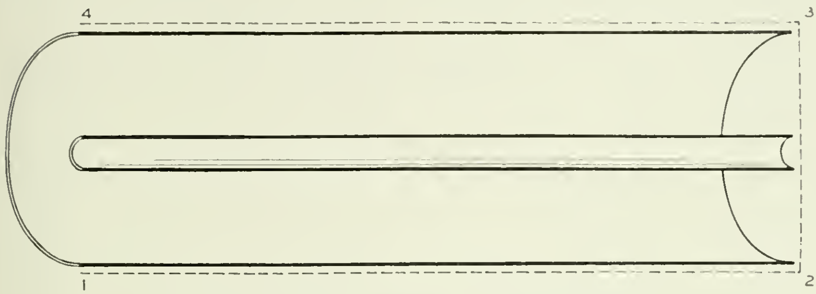


Fig. 7*a*—An axial cross-section of a coaxial pair.

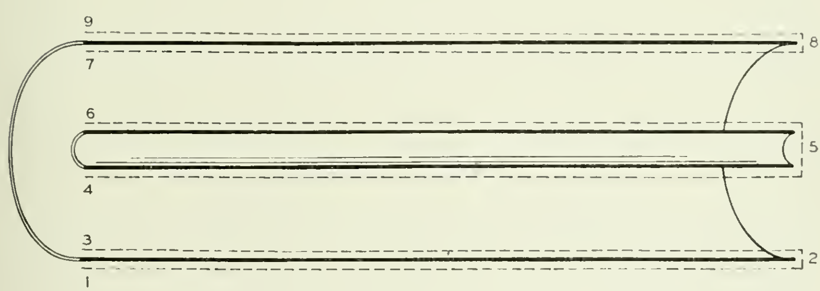


Fig. 7*b*—An axial cross-section of a coaxial pair.

are small by comparison with the wave-length. We shall assume that the coaxial pair is semi-infinite. Trusting his common sense, the engineer assumes that inside this structure the magnetic lines are circles coaxial with the cylinders. The electric lines are the radii and the electric current in the cylinders as well as the transverse voltage between the cylinders vary along the length in the same way as in a transmission line with uniformly distributed series inductance and shunt capacity. A careful analysis by John R. Carson indicates that this simple picture is justifiable if the cross-section of the coaxial pair is small by comparison with the wave-length.[10] While a whole series of electric waves can exist in such a structure, all of these waves except the one recognized by the engineer, the *principal wave*, are attenuated very rapidly and are significant only very close to the generator and

[10] John R. Carson, "The Guided and Radiated Energy in Wire Transmission," *A. I. E. E. Journal*, pp. 908–913, October, 1924.

very near the open end. The complementary waves are needed only for logical consistency and to satisfy the boundary conditions.

Thus let us suppose that the field distribution in the coaxial pair is known to a high degree of accuracy. In order to calculate the field outside the coaxial pair and hence obtain the radiated power we can use the Equivalence Principle in two ways. We can fit our surface C smoothly over the outer cylinder and the open end (Fig. 7*a*) or, regarding this surface as a perfectly elastic rubber sheet, we can press it through the open end and fit it smoothly over the inner surface of the outer conductor and the outer surface of the inner conductor (Fig. 7*b*). Since by hypothesis the conductors are perfect, the components of E tangential to the cylinders vanish; hence in the second choice of C the equivalent layer consists of only an electric current sheet. Naturally this current distribution is precisely that which actually exists in the conductors so that this choice of C leads to something that we knew beforehand, namely: if the actual sources, that is, if the electric currents in the structure are known exactly or approximately, the entire field can be calculated exactly or approximately.

The first choice of C is more important. Over the lateral portion (12, 34) of C the equivalent magnetic current sheet vanishes as in the preceding case on account of the perfect conductivity of the cylinders. The magnetic intensity just outside the coaxial pair is also zero except near the open end where it must be exceedingly small. To see this, we need only recall that the electric currents in the two cylinders are equal and opposite and that except in the neighborhood of the open end the displacement currents are transverse. Thus the equivalent electric current sheet can be ignored altogether. What is left is the magnetic current sheet over the surface of the open end; the density of this sheet is determined by the radial component of the electric intensity and in the final analysis by the voltage existing between the ends of the inner and outer conductors. Presently we shall carry out the actual calculations but just now we shall examine the question of the accuracy of the results. Of course, the results would be exact if we knew the equivalent electric and magnetic sheets accurately; and the above approximations appear to be reasonable. We shall not be able to find out how good these approximations are but we can prove that they are just as good as the approximations usually made in calculating the radiated power from the distribution of electric currents. The only virtue of the Equivalence Principle is to save a certain amount of mathematical work and furnish a further insight into the phenomena of radiation.

If a progressive wave is advancing from left to right in a semi-infinite coaxial pair (Fig. 7) and if the generator is at infinity, we can assume it to be the principal wave. At the open end this wave is reflected. It is usually assumed that the reflected wave is also the principal wave but moving in the opposite direction. In other words, it is assumed that the total field is such that the electric lines are radial and the magnetic lines are circular. Since the electric lines are radial, there is no longitudinal displacement current; and since the conduction current at the open end must be zero, the magnetic intensity is zero over the entire open end. This is what follows if we neglect the complementary waves.

These approximate results correspond to the *exact* results in the following hypothetical situation. If a hypothetical perfect magnetic conductor is fitted over the open end of the coaxial pair so that it closes it entirely, then the reflection is complete and there are no complementary waves. Perfect magnetic conductors are defined by analogy with perfect electric conductors—the tangential component of the magnetic intensity vanishes at the surface of the former just as the tangential component of the electric intensity vanishes at the surface of the latter. Magnetic conductors support magnetic current sheets just as electric conductors support electric current sheets. The densities of the sheets are given by the discontinuities of the tangential components of E in the former case and H in the latter.

In the hypothetical case in which the open end is closed with a perfect magnetic conductor, no energy can flow outside the coaxial pair. This is because the flow of energy is given by $\frac{1}{2}E \times H$ and either one or the other factor vanishes over the outer boundary of the structure. The field outside the coaxial pair must now be identically zero. Our sources are the electric current in the walls of the coaxial pair and the magnetic currents in the cap. If one field is designated by S and the other by S', then $S + S' = 0$ and $S' = -S$. Thus the field produced by the electric currents in the conductors on a supposition that principal waves alone exist, is the same as the field produced by a hypothetical distribution of magnetic currents over the surface of the open end.

Let us examine another case. It is usual to assume that the electric current in a thin wire (Fig. 8) in free space is distributed sinusoidally. Experimental evidence shows that the radiated power calculated on this assumption is very nearly correct. On the other hand the sinusoidal distribution of the electric current in the wire corresponds to a hypothetical case in which a perfect magnetic conductor is introduced

in the shape of a sphere concentric with the center of the wire and passing through its ends. Thus we could calculate the radiated power from an appropriate distribution of magnetic currents over this sphere but in this case such a procedure would involve more difficult integrations than the usual method.

Before considering the more general case of radiation from a semi-infinite coaxial pair let us assume that the radii of the two conductors are nearly equal. We have seen that in applying the Equivalence Principle we need take into account only the magnetic current sheet over the open end of the pair. In the present instance this sheet is merely a circular loop of magnetic current equal to the voltage V between the ends of the conductors. If we were to treat in the same

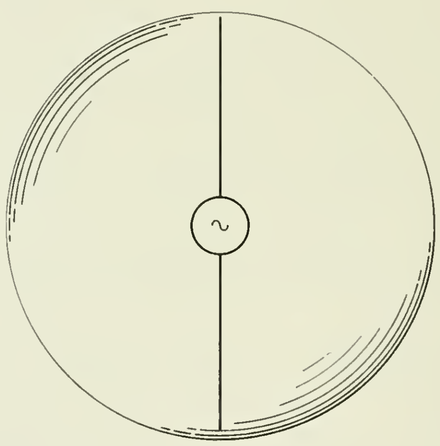


Fig. 8—A vertical antenna and a cross-section of an imaginary sphere passing through the ends of the antenna.

manner a condenser made of two parallel circular plates, we should come to the conclusion that it is also equivalent to a magnetic loop around its periphery. Thus in both cases the radiated power is the same. But the power radiated by an electric doublet is known to be $(40\pi^2 I^2 l^2)/\lambda^2$ watts where I is the amplitude of the electric current, l the length of the doublet and λ the wave-length. In applying this formula to a condenser it is better to express it in terms of the voltage V across the condenser and its area S. The capacity of the condenser is $C = S/(36\pi 10^{11} l)$ farads and $I = \omega CV = SV/60\lambda l$ amperes. Hence the power radiated by the condenser is $(\pi^2 S^2 V^2)/90\lambda^4$ watts. This is also the approximate power radiated by the coaxial pair if we interpret S as the cross-sectional area of either conductor.

Let us now calculate the more general expression for the power radiated from an open end of a coaxial pair. The cylindrical conductors whose cross-sections are shown in Fig. 9 are supposed to extend

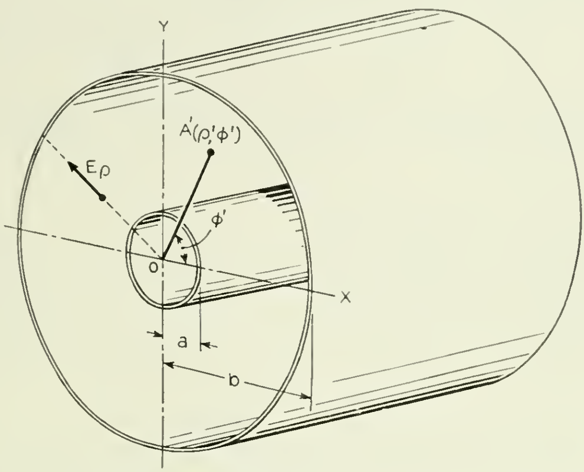


Fig. 9—The end view of a coaxial pair of cylindrical conductors.

below the surface of the paper and the z-axis of the coordinate system is directed toward the reader. The primed letters will refer to points situated in the opening, the unprimed letters being reserved for typical points in space.

The electric intensity in the coaxial pair varies inversely as the distance from the axis

$$E_{\rho'} = \frac{P}{\rho'}, \qquad E_{\varphi'} = 0. \tag{21}$$

In accordance with the Equivalence Principle we assume that the field below the xy-plane is wiped out and the discontinuity in E arising as the result of the separation is replaced by a magnetic current sheet. This magnetic current is perpendicular to E; by (14) its density is

$$M_{\varphi'} = -E_{\rho'} = -\frac{P}{\rho'}, \qquad M_{\rho'} = 0. \tag{22}$$

The constant E is related to the voltage V between the ends of the coaxial conductors; in fact, we have

$$V = \int_a^b E_{\rho'} d\rho' = P \log \frac{b}{a}, \qquad P = \frac{V}{\log \frac{b}{a}} \cdot \tag{23}$$

In order to calculate the field at some point A due to the distribution (20), we must determine the retarded electric vector potential. Since the integration is vectorial, it is convenient to deal with the cartesian components of the magnetic current density

$$M_{x'} = \frac{P \sin \varphi'}{\rho'}, \qquad M_{y'} = -\frac{P \cos \varphi'}{\rho'}. \tag{24}$$

The area of the element is $\rho' d\rho' d\varphi'$ so that the components of the retarded potential are

$$\begin{aligned} F_x &= \frac{1}{4\pi} \int_a^b \int_0^{2\pi} \frac{M_{x'} e^{-i\beta AA'} \rho' d\rho' d\varphi'}{AA'} \\ &= \frac{P}{4\pi} \int_a^b \int_0^{2\pi} \frac{e^{-i\beta AA'} \sin \varphi'}{AA'} d\rho' d\varphi', \\ F_y &= -\frac{P}{4\pi} \int_a^b \int_0^{2\pi} \frac{e^{-i\beta AA'} \cos \varphi'}{AA'} d\rho' d\varphi', \quad \beta = \omega\sqrt{\mu\epsilon} = \frac{\omega}{c} = \frac{2\pi}{\lambda}. \end{aligned} \tag{25}$$

Hence the components in the polar coordinates are

$$\begin{aligned} F_\varphi &= -F_x \sin \varphi + F_y \cos \varphi \\ &= -\frac{P}{4\pi} \int_a^b \int_0^{2\pi} \frac{e^{-i\beta AA'} \cos (\varphi - \varphi')}{AA'} d\rho' d\varphi', \\ F_\rho &= 0. \end{aligned} \tag{26}$$

The distance AA' is

$$AA' = \sqrt{r^2 - 2r\rho' \cos \vartheta + \rho'^2}, \tag{27}$$

where r is the distance OA, and ϑ is the angle between OA and OA'. Since we are interested in the radiation field alone, we need retain only those terms in (24) which vary inversely as the distance; the other terms contribute nothing to the radiated power. Thus we let r increase indefinitely, obtaining

$$AA' = r - \rho' \cos \vartheta, \tag{28}$$

and then

$$F_\varphi = -\frac{Pe^{-i\beta r}}{4\pi r} \int_a^b \int_0^{2\pi} e^{i\beta\rho' \cos \vartheta} \cos (\varphi - \varphi') d\rho' d\varphi'. \tag{29}$$

If θ and θ' are the angles made by OA and OA' with OZ, we have

$$\begin{aligned} \cos \vartheta &= \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos (\varphi - \varphi') \\ &= \sin \theta \cos (\varphi - \varphi'). \end{aligned} \tag{30}$$

Since ρ' is small by comparison with the wave-length λ, we can expand the exponential term in the integrand into a power series and retain

only the first two terms

$$e^{i\beta\rho'\cos\vartheta} \doteq 1 + i\beta\rho'\cos\vartheta = 1 + i\beta\rho'\sin\theta\cos(\varphi - \varphi'). \qquad (31)$$

We need the second term because the integral of the first vanishes. Integrating the second term, we obtain

$$\begin{aligned} F_\varphi &= -\frac{i\beta P e^{-i\beta r}\sin\theta}{4\pi r}\int_a^b \rho' d\rho' \int_0^{2\pi}\cos^2(\varphi - \varphi')d\varphi' \\ &= -\frac{1}{8} i\beta(b^2 - a^2)P\,\frac{e^{-i\beta r}}{r}\sin\theta. \end{aligned} \qquad (32)$$

The magnetic current is uniform around the axis and there is no accumulation of magnetic charge anywhere; hence the second term in the expression for H as given by (11) vanishes. Therefore

$$H_\varphi = -\frac{P}{8}\omega\epsilon\beta(b^2 - a^2)\frac{e^{-i\beta r}}{r}\sin\theta. \qquad (33)$$

At a great distance from the source the wave tends to become plane so that in the radiation field the electric intensity is perpendicular to OA and to H and is given by

$$E_\theta = \sqrt{\frac{\mu}{\epsilon}}H_\varphi = 120\pi H_\varphi. \qquad (34)$$

According to Poynting the radiated power is the real part of the following integral

$$W = \frac{1}{2}\int_0^\pi\int_0^{2\pi} E_\theta H_\varphi^* r^2 \sin\theta\, d\theta\, d\varphi \text{ watts}, \qquad (35)$$

where H_φ^* is the complex number conjugate to H_φ. Substituting from (31) and (32), we obtain

$$\begin{aligned} W &= \frac{\pi^3}{980}\left(\frac{b^2 - a^2}{\lambda^2}\right)^2 P^2 \int_0^\pi \sin^3\theta\, d\theta \int_0^{2\pi} d\varphi \\ &= \frac{\pi^4}{360}\left(\frac{b^2 - a^2}{\lambda^2}\right)^2 P^2 \text{ watts}. \end{aligned} \qquad (36)$$

Introducing from (21) the expression for P and designating by S the area of the opening, we have

$$W = \frac{\pi^2}{360}\left(\frac{S}{\lambda^2 \log\frac{b}{a}}\right)^2 V^2 \text{ watts}. \qquad (37)$$

The effect of radiation on the transmission line can be simulated by a resistance R,

$$R = \frac{180}{\pi^2}\left(\frac{\lambda^2 \log \frac{b}{a}}{S}\right)^2 \text{ ohms} \tag{38}$$

shunted across the open end. This is not the resistance seen by the generator. If V and I are the amplitudes of the voltage and the electric current at their antinodes and Z_0 the characteristic impedance of the coaxial pair, then

$$V = Z_0 I = \left(60 \log \frac{b}{a}\right) I. \tag{39}$$

Since the end of the coaxial pair is a voltage antinode, the radiated power may be expressed as

$$W = 10\pi^2 \left(\frac{S}{\lambda^2}\right)^2 I^2 \text{ watts.} \tag{40}$$

Hence the radiation resistance seen by a generator placed at a current antinode is

$$R_G = \frac{20\pi^2 S^2}{\lambda^4} \text{ ohms.} \tag{41}$$

With this simple illustration, we conclude the present paper.

Magnetic Alloys of Iron, Nickel, and Cobalt *

By G. W. ELMEN

The unexpected magnetic properties of certain alloys of iron and nickel discovered some 20 years ago led to a thorough study of the entire range of iron-nickel alloys. The results of this study were so encouraging that alloys of these metals with cobalt, the only other ferromagnetic metal, also were studied, as well as various alloys of these metals with small amounts of non-magnetic metals added. From the results of this extended investigation have emerged several alloys that are playing important parts in the continued advancement of electrical communication.

SOME alloys of iron, nickel, and cobalt have remarkable magnetic properties superior in many situations to those of the constituent metals. Many of these alloys have found wide use in the instrumentalities and circuits of electrical communication, and were developed primarily for that purpose. This paper reports the experience and techniques of the Bell Telephone System in the development and utilization of these materials.

The advantageous properties of these alloys were disclosed through exhaustive researches, during which the whole realm of combinations of these three metals was explored. That certain alloys of iron and nickel had unexpected properties at low flux densities had already been discovered in the Bell Telephone Laboratories. There was at that time no theoretical basis for predicting, or even explaining, the character of those alloys; and, therefore, a study was undertaken of the whole iron-nickel series. The results were so encouraging that combinations of these elements with cobalt likewise were studied; and finally those alloys of special interest were combined with varying amounts of non-magnetic metals. In the course of this investigation several thousand specimens were made and tested in a period extending over fifteen years.

Such an empirical investigation is time consuming and expensive, but in a field where so little theory was available for guidance it was the only certain means to determine the practical possibilities of these alloys. It has been justified by the large number of alloys it has developed for practical use in communication engineering. One of the first and most striking applications was to submarine telegraph cables. The largest field of application, however, has been in teleph-

* Published in the December 1935 issue of *Electrical Engineering*, and scheduled for presentation at the Winter Convention of the A. I. E. E., New York City, January 28–31, 1936.

ony, where the requirements generally are very exacting, and where other advances have imposed rigid demands on the magnetic materials.

In telephone circuits, standards of transmission efficiency require that the magnetic materials used as circuit elements shall produce maximum magnetic effect with minimum energy loss and distortion of the transmitted currents. Translated into magnetic characteristics, this means that at low magnetizing forces the material shall have high permeability in combination with low hysteresis loss, and, in many situations, constancy of permeability over the operating range. In circuits for voice and carrier currents it is often necessary to reduce the intrinsic permeability of the material to obtain the required constancy and low losses in the apparatus. Furthermore, to minimize eddy currents, a high resistivity is required and the material must be structurally suitable for fabrication into thin laminations. For other uses, such as for signaling and switching mechanisms, the magnetic properties at medium and high flux densities determine the suitability of the material. High permeability and low coercive force make for improved sensitivity and speed of operation. Low coercive force is of special interest in marginal apparatus where the difference between the operating and releasing currents is small.

Preparation and Composition of the Alloys

A great many factors contribute to the final properties of an alloy. Among the most important of these are the purity of the elements used in the alloy, their preparation, and the heat-treatment. The magnetic properties attainable can be completely masked by the intrusion of small quantities of certain impurities or by improper heat-treatment. For iron the magnetic properties can be improved materially by removing extremely small quantities of carbon and other non-metallic elements through heat-treating * in an atmosphere of hydrogen and at temperatures close to the melting point. This method of purification also improves the magnetic properties for alloys of iron, nickel, and cobalt. For communication purposes, it has not been found expedient as yet to introduce this method of refinement in the commercial production of these alloys. The purity of the constituents is controlled by ordinary methods of chemical analysis, by methods of melting, and by annealing processes which do not increase the amounts of important

* There is a rapidly growing technical literature relating to the effects of very small percentages of impurities on magnetic properties and the methods for their removal, with notable contributions by T. D. Yensen of the Westinghouse Electric and Manufacturing Company, W. E. Ruder of the General Electric Company, and P. P. Cioffi of the Bell Telephone Laboratories.

impurities. The magnetic properties recorded in this paper have therefore been confined mostly to those obtained on materials produced by standard metallurgical methods.

In the commercial method of producing these alloys the best grades of commercial iron, nickel, and cobalt are used. The melting is done in an electric furnace, and after the mechanical fabrication into suitable shapes these alloys are heat-treated to develop the desired magnetic properties.

Early in an investigation of these alloys it was found that some of them required special heat-treatments to develop the desired magnetic properties. For some the slow cooling incident to the ordinary process of annealing was not suitable, and a rapid cooling was necessary. For another group the slow cooling in the annealing process was not slow enough, and the best results were obtained when the alloys were held at a constant high temperature for a considerable time. It was evident that to determine the most suitable temperature of heating and rate of cooling for each alloy would require more time than was warranted in the exploratory work. Three methods of heat-treatment that, in a general way, would separate the alloys into groups, were developed. These heat-treatments are designated in this paper as "annealing," "quenching," and "baking."

The annealing process consists of heating the samples in closed containers to a temperature of 1,000 degrees centigrade, and cooling with the furnace. The cooling ordinarily requires 7 hours before room temperature is reached. This heat-treatment is primarily for the purpose of removing the effects of mechanical strains necessarily resulting from the rolling and stamping of the alloys into suitable shapes. All the alloys discussed in this paper received this heat-treatment before any of the more special processes were applied.

The quenching process consists of heating the alloys for a short time at 600 degrees centigrade, and cooling in air at room temperature for small samples with large surfaces, and in oil for larger samples. The rate of cooling attained by these methods is approximately 40 degrees centigrade per second. It has been found that the best rate of cooling for maximum permeability does not always develop the highest initial permeability. The difference, however, is not large, and often is masked by other variations in the manufacturing process.

The baking process consists of heating the alloys for 24 hours at 425 degrees centigrade, and then slowly cooling to room temperature. The rate of cooling does not affect the development of the magnetic properties unless it is so rapid as to introduce mechanical strains.

Classification of the Alloys

A convenient way of showing graphically the compositions of the alloys of iron, nickel, and cobalt is by means of the composition triangle in Fig. 1. The sides of this triangle represent the binary alloys of the three metals, and points inside the triangle, the ternary alloys.

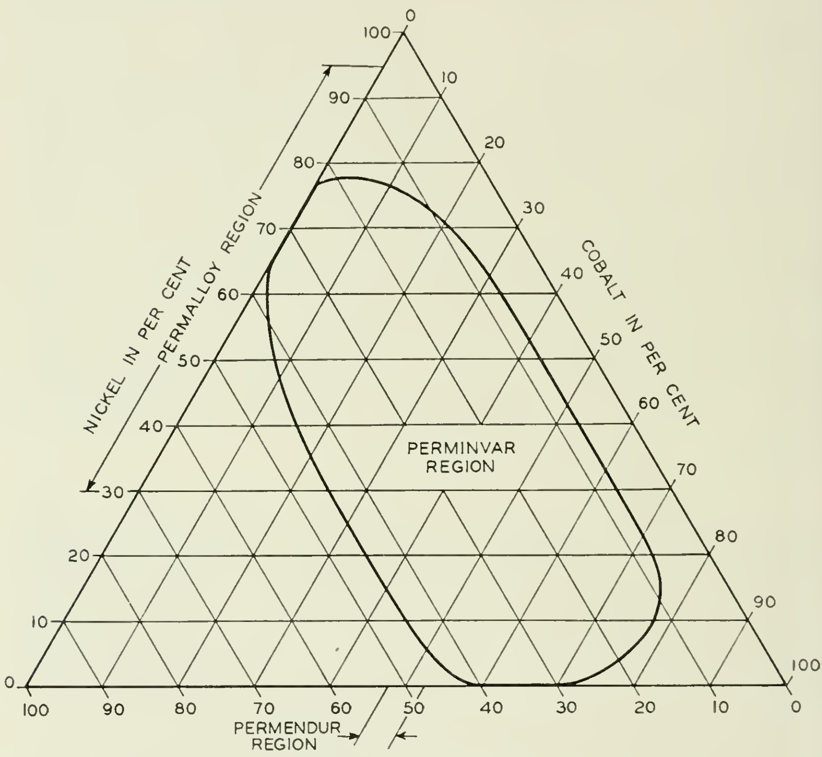


Fig. 1—Composition diagram for alloys of iron, nickel and cobalt.

In this diagram the alloys of special interest because of their magnetic properties are indicated, and, for convenience, each group in which the magnetic properties are similar has been given a specific name.

On the iron-nickel side of the triangle the permalloy region is indicated. In this group several compositions have been developed for commercial use in the Bell System. The method of identifying these alloys consists of prefixing a numeral indicating the per cent of nickel; for example, 45-permalloy contains 45 per cent nickel and 55 per cent iron. To some of these permalloys small amounts of other metals also

are added. In designating ternary permalloys the same scheme is extended, so that the name gives everything except the iron content, and this is obtainable by difference. Thus, 3.8–78.5 Cr-permalloy contains 3.8 per cent chromium, 78.5 per cent nickel, and 17.7 per cent iron.

The perminvar region, enclosed by the curved line, contains those compositions that require baking to develop completely their characteristic magnetic properties. The specific compositions of these alloys are indicated by two prefixed numerals, the first indicating the nickel and the second the cobalt percentages, respectively. Thus the 45–25 perminvar alloy contains 45 per cent nickel, 25 per cent cobalt, and 30 per cent iron. Another alloy of the perminvar group, in which the nickel and cobalt percentages are the same as the alloy just mentioned, but which contains 7 per cent molybdenum and 23 per cent iron, is designated as 7–45–25 Mo-perminvar.

In the iron-cobalt series of alloys the composition 50 per cent iron and 50 per cent cobalt has been developed for commercial use. This is the permendur alloy, indicated in the triangular diagram in Fig. 1. This alloy is difficult to cold roll, but the addition of 1.7 per cent vanadium improves the mechanical properties and makes it sufficiently ductile to roll into thin sheets. The same system has been followed in designating this alloy as in the case of the permalloys. Thus 1.7 V-permendur is an alloy containing 1.7 per cent vanadium with iron and cobalt in equal proportions.

Table I lists the designations and compositions of those alloys, developed for particular purposes, which are discussed more fully in the remainder of this paper.

TABLE I

Designations and Compositions of Some Magnetic Alloys

Designation	Composition, Per Cent					
	Ni	Fe	Co	Cr	Mo	V
78.5 permalloy	78.5	21.5				
80 permalloy	80	20				
45 permalloy	45	55				
3.8–78.5 Cr-permalloy	78.5	17.7		3.8		
3.8–78.5 Mo-permalloy	78.5	17.7			3.8	
2–80 Mo-permalloy	80	18			2	
45–25 perminvar	45	30	25			
7–45–25 Mo-perminvar	45	23	25		7	
Permendur		50	50			
1.7 V-permendur		49.15	49.15			1.7

Ni = nickel; Fe = iron; Co = cobalt; Cr = chromium; Mo = molybdenum; V = vanadium.

In Figs. 2, 3, and 4 are shown the magnetization curves for low, medium, and high magnetizing forces for these alloys, except the 80 permalloy and the 1.7 V-permendur, for which the curves are substantially the same as for 78.5 permalloy and permendur, respectively. Curves for "Armco" iron and ordinary commercial 4 per cent silicon steel also are shown in these figures. All these materials were annealed, and in the case of the 78.5 permalloy and the perminvars the annealing was followed by quenching and baking, respectively. It

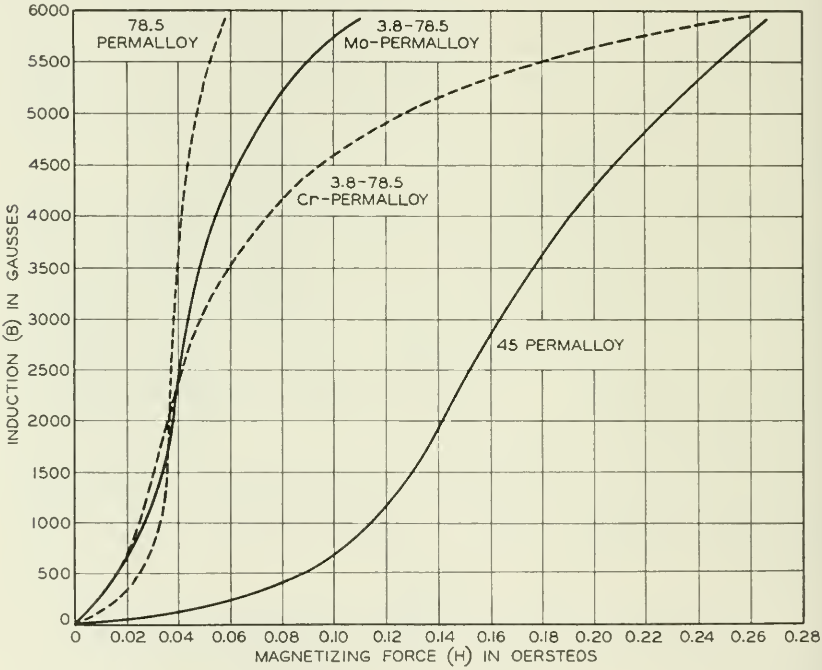


Fig. 2—Magnetization curves for several permalloys for flux densities less than 6,000 gausses.

may be seen from these curves that the permalloy group reaches almost saturation values long before the iron and silicon steel and the other alloys have reached the lower bend in the magnetization curve. With the exception of the 45-permalloy, which saturates at a fairly high value, the permalloys have low saturation induction and the permendur the highest. The permeability curves computed from these curves are plotted in Figs. 5 and 6. In Fig. 5, curves for the permalloy alloys are plotted at a smaller vertical scale than in Fig. 6 containing the curves for the other alloys.

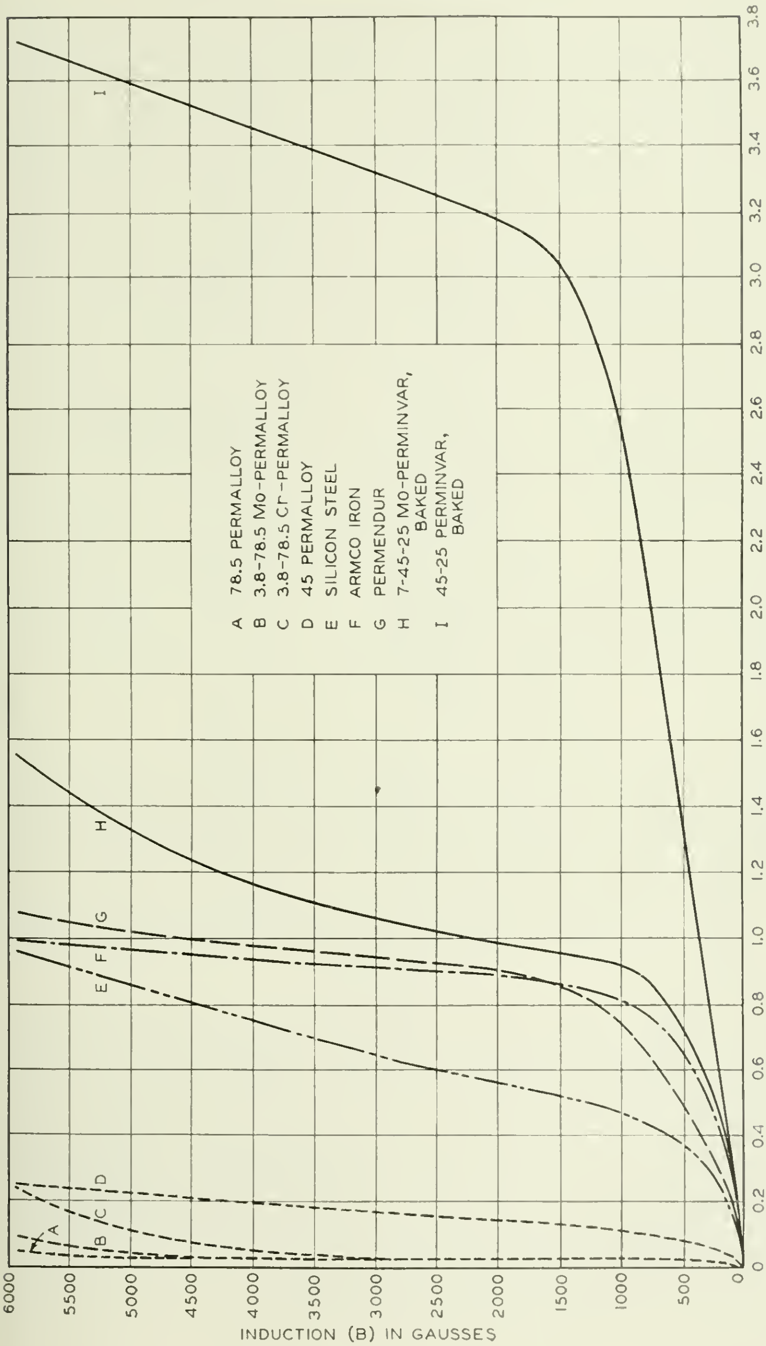


Fig. 3—Magnetization curves for permalloys, perminvars, and permendur.

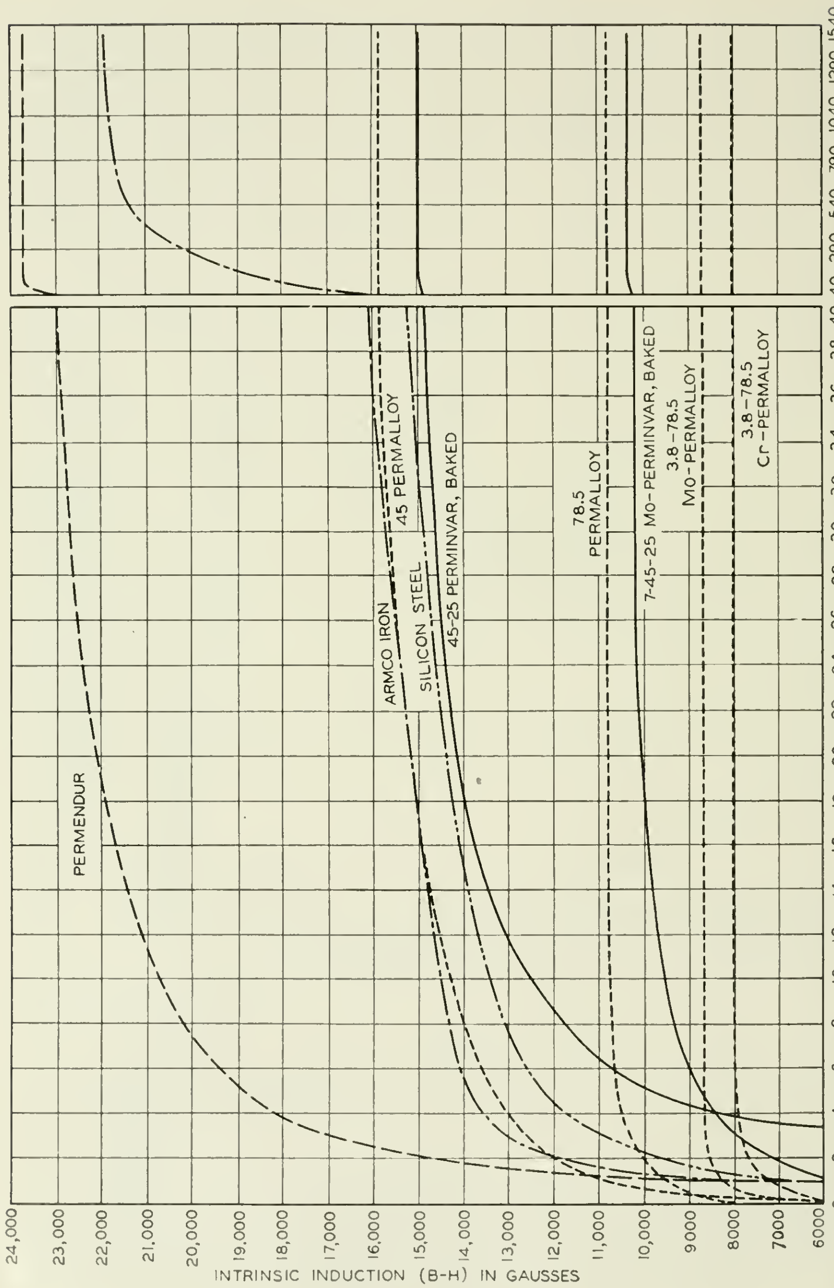


Fig. 4—Magnetization curves for permalloys, perminvars, and permendur.

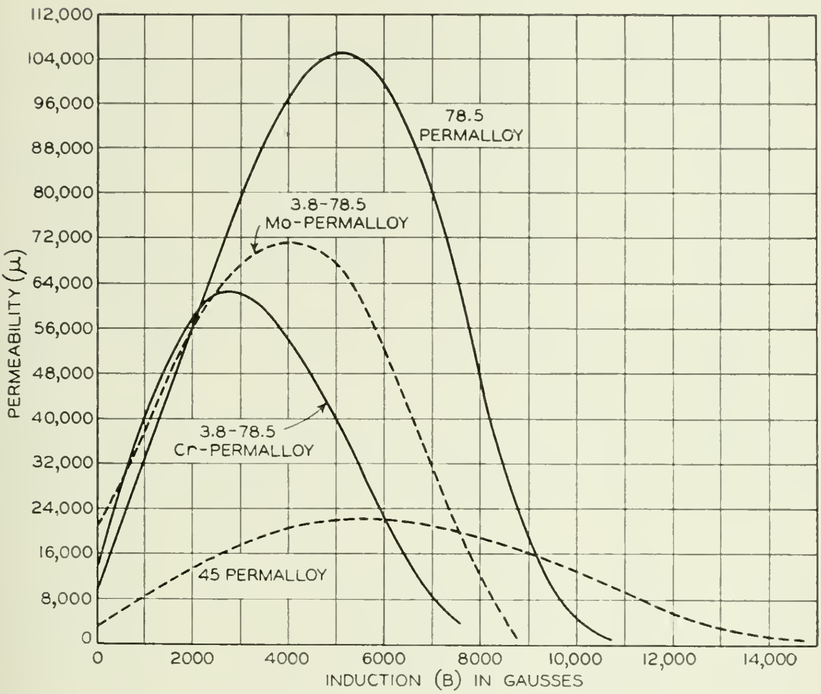


Fig. 5—Permeability curves for permalloys.

The permeability for alternating current of small constant amplitude as a function of superposed d.-c. magnetizing force is shown in Fig. 7 for some of the alloys. In most apparatus where both alternating and direct current are involved, this "butterfly curve" must be relatively flat over the expected range of d.-c. excitation. The important magnetic constants for these alloys are given in table II.

Permalloys

In Fig. 8 the initial and maximum permeabilities and the coercive force and resistivities are plotted for quenched alloys of the iron-nickel series. These curves show the remarkable variations in magnetic properties with composition in this series of alloys. The permalloy region includes alloys between 30 and 95 per cent nickel, as indicated in Fig. 1. Some of the alloys in this region, particularly from 50 to 85 per cent nickel, require rapid cooling to develop the magnetic properties indicated in the curves. If they are merely annealed, both the maximum and the initial permeabilities are much lower. The greatest effect in reducing the permeabilities by slow cooling appears to be for

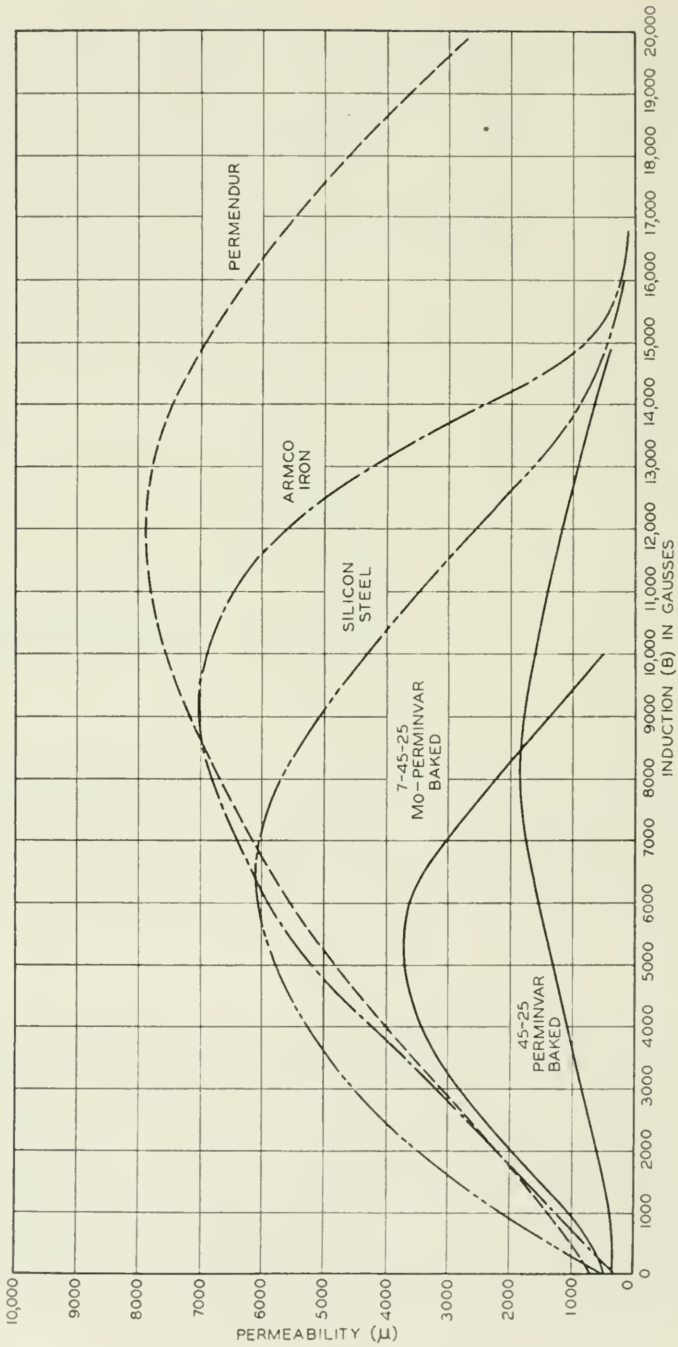


Fig. 6—Permeability curves for perminvars and permendur.

TABLE II

MAGNETIC CONSTANTS FOR ALLOYS DISCUSSED IN THIS PAPER

Material	μ_0	μ_m	$W_{H=\infty}$	B_r	H_c	$(B-H)_{H=\infty}$	ρ
"Armco" iron	250	7,000	5,000	13,000	1.0	22,000	11
4% silicon-steel	600	6,000	3,500	12,000	0.5	20,000	50
78.5 permalloy, quenched	10,000	105,000	200	6,000	0.05	10,700	16
45 permalloy	2,700	23,000	1,200	8,000	0.3	16,000	45
3.8–78.5 Cr-permalloy	12,000	62,000	200	4,500	0.05	8,000	65
3.8–78.5 Mo-permalloy	20,000	75,000	200	5,000	0.05	8,500	55
45–25 perminvar, baked	400	2,000	2,500	3,000	1.2	15,500	19
7–45–25 Mo-perminvar, baked	550	3,700	2,600	4,300	0.65	10,300	80
Permendur	700	7,900	6,000	14,000	1.0	24,000	6

Here μ_0 and μ_m are the initial and maximum permeabilities, respectively; $W_{H=\infty}$ is the hysteresis loss in ergs per cubic centimeter per cycle for saturation value of flux density; B_r is the residual induction in gausses; H_c is the coercive force in oersteds; $(B - H)_{H=\infty}$ is the saturation value of the intrinsic induction in gausses; ρ is the resistivity in microhms-centimeter.

the alloys containing between 70 and 80 per cent nickel; for example, 78.5 permalloy with a standard anneal has its initial permeability reduced to 1,200. If the alloy is baked for several hundred hours this permeability can be reduced still further to about 500. There is a very rapid decrease in the coercive force as the nickel increases above 27 per cent, and the lowest values are reached in the region between 70 and 80 per cent nickel. The resistivity increases rapidly just below the permalloy region, and reaches maximum at about 31 per cent nickel. It should be noted that the large changes in the coercive force and the resistivity are at the lower end of the permalloy region, while the highest permeabilities are developed in the alloys containing between 75 and 80 per cent nickel.

45-Permalloy

One of the alloys developed for commercial use is 45-permalloy. This attains a saturation flux density as high as any of the permalloys. At 40 oersteds the flux density is 16,000 gausses, substantially the same as for "Armco" iron, and considerably higher than for ordinary silicon steel (Fig. 4). The initial and maximum permeabilities (under standard practice of heat-treating) are 2,700 and 23,000, respectively (Fig. 5 and Table II). For cores requiring flux densities between 8,000 and 12,000 gausses this alloy is specially useful. The resistivity of the alloy is 45 microhms-centimeter, which is high enough to make it superior for use in cores in a.-c. circuits. The higher permeability at

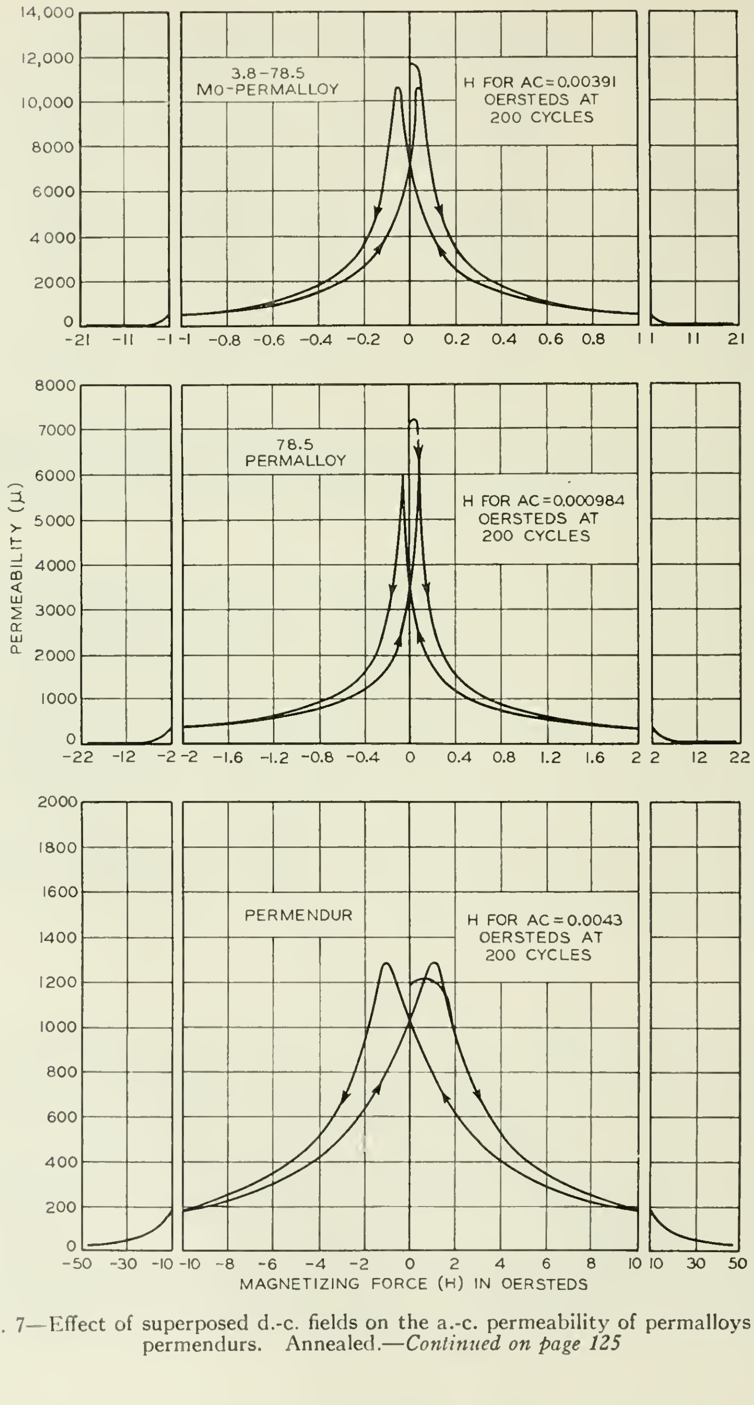


Fig. 7—Effect of superposed d.-c. fields on the a.-c. permeability of permalloys and permendurs. Annealed.—*Continued on page 125*

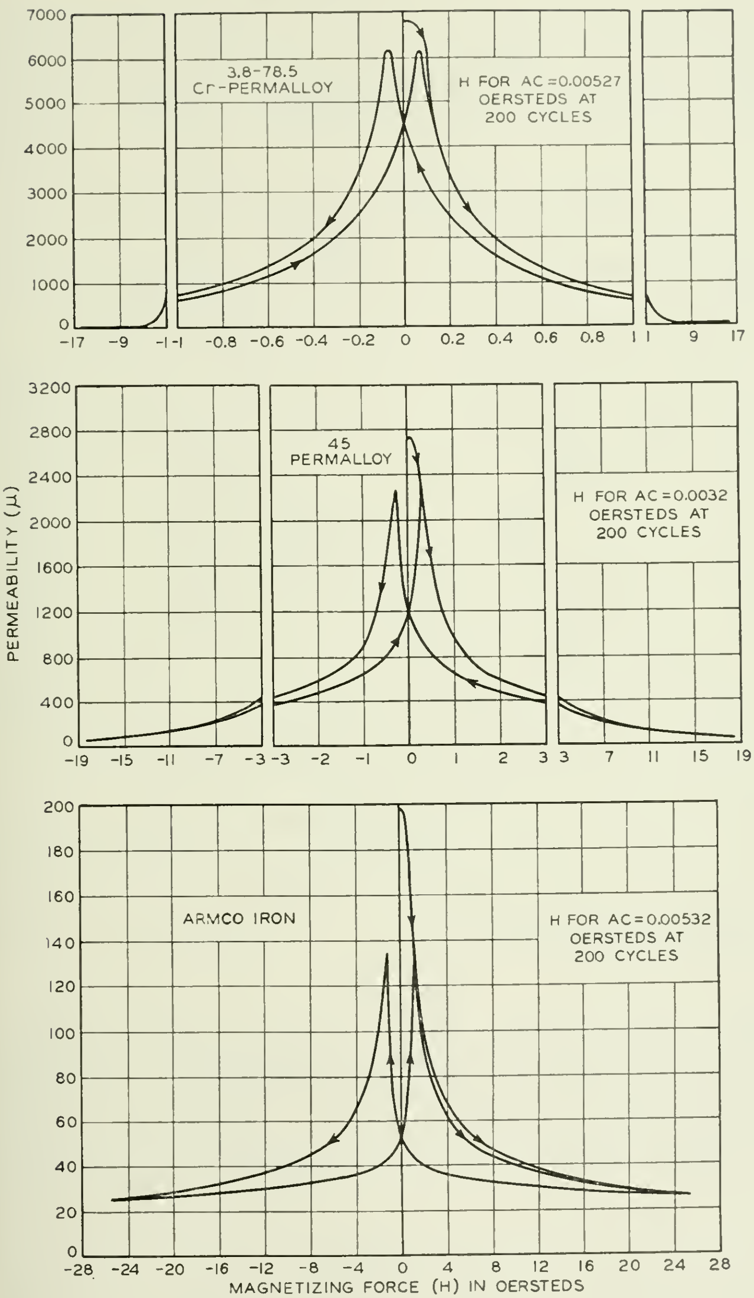


FIG. 7—*Continued from page 124*

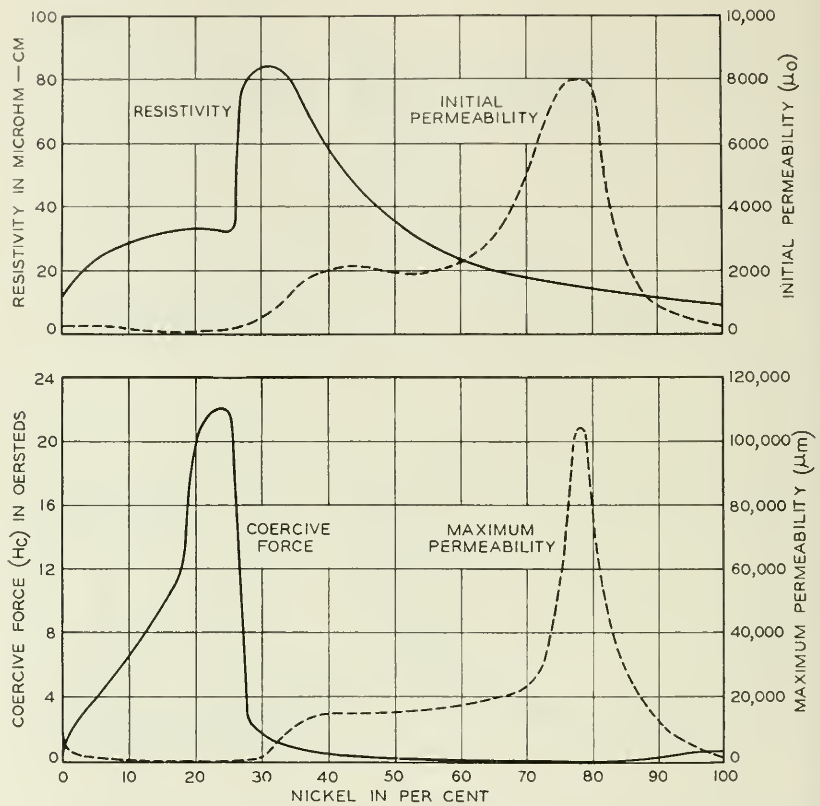


Fig. 8—Resistivity, initial and maximum permeabilities, and coercive force of iron-nickel alloys.

fairly high values of superposed d.-c. field, shown in Fig. 7, also favors its use for some purposes.

78.5-Permalloy

Another alloy long used in the telephone plant is 78.5 permalloy. Quenching develops a higher maximum permeability in this than in any of the other permalloys. Initial and maximum permeabilities of 10,000 and 105,000 readily are developed. The hysteresis loss and the coercive force of quenched 78.5 permalloy are minimum. The saturation flux density of this alloy is between 10,000 and 11,000 gausses, and it is reached with a very low magnetizing force. The rapid rise in the flux density of this alloy for small increments in the magnetizing force and the low saturation flux density are shown in Figs. 2, 3, and 4.

Initial and maximum permeabilities of 78.5 permalloy are improved by elimination of impurities and also by special care in the quenching process. As stated earlier, the rates of cooling required to develop the highest initial permeability differ from those for the highest maximum.

Chromium Permalloy and Molybdenum Permalloy

When other metals are added to permalloys their resistivities, in general, are increased. In research work at the Bell laboratories chromium and molybdenum mostly were used. It was found that with these elements a desirable combination of high resistivity and high initial permeability could be obtained. The variation in resistivity, keeping the nickel content constant at 78.5 per cent, is shown in Fig. 9. Chromium increases the resistivities somewhat more than molybdenum for a given addition, but the difference is not very large. The 3.8–78.5 Cr-permalloy has a resistivity of 65 microhms-centimeter, as compared with 55 for the 3.8–78.5 Mo-permalloy.

Figure 9 also illustrates the manner in which additions of these metals affect the initial permeability and the sensitivity of the permeability to rate of cooling. The solid line curves are for the annealed and the broken-line curves for the quenched specimens. For the quenched alloys the highest permeabilities are obtained when the added chromium and molybdenum are 2.4 per cent and 1.6 per cent, respectively. For this cooling rate the chromium permalloy seems to develop a slightly higher initial permeability. The difference, however, is small, and a greater spread between different samples has been observed. For the annealed alloys the largest value of initial permeability is obtained with molybdenum permalloy. For 3.8 Mo-permalloy an initial permeability of 20,000 is obtained. With the same heat-treatment the initial permeability of the corresponding chromium alloy is 12,000. It is surprising to note that small additions of these non-magnetic metals increase the initial permeability to values considerably higher than that for quenched 78.5 permalloy. Beyond 5 per cent this improvement ceases. All additions decrease the saturation induction values and the maximum permeabilities.

Several of these alloys have been developed for commercial use. Of these the most important are 2–80 Cr-permalloy, 3.8–78.5 Cr-permalloy, and 3.8–78.5 Mo-permalloy.

Perminvar

The distinctive magnetic properties of the perminvars are constancy of permeability at low flux densities, a low hysteresis loss in the same

Fig. 9—Resistivity and initial permeability of 78.5 permalloy with chromium or molybdenum replacing part of the iron. In the curve designations, X indicates the percentage of chromium (Cr) or molybdenum (Mo) as read from the abscissa.

range, and, for medium flux densities, a characteristic constriction in the middle of the hysteresis loop. In some alloys this constriction is so extreme that the coercive force vanishes, making the two branches of the loop coincide when the magnetizing force is reduced to zero, in spite of the considerable hysteresis loss involved in the entire cycle. At high flux densities this constriction disappears and the loops have normal shapes.

The degree to which these properties can be developed depends on the composition and the heat-treatment. For the most typical alloys slow cooling in the annealing process produces this effect to a certain degree. Baking for 24 hours in the 400–500 degree (centigrade) temperature range brings most alloys into a stable condition in which no further baking materially will affect the magnetic properties.

As indicated in Fig. 1, some of the binary alloys tend toward the perminvar characteristics with long baking. Of the permalloys a considerable proportion of those that must be quenched to develop the desirable magnetic properties show perminvar characteristics when they are baked.

45–25 Perminvar

The perminvar characteristics have been developed most intensely in 45–25 perminvar. The magnetization curve in Fig. 3, and the permeability curve in Fig. 6, illustrate this fact. The constancy of permeability at low magnetizing forces and the necessity of "baking" to attain this condition are illustrated in one of the sections of Fig. 10, where the permeabilities are plotted for the quenched and baked conditions. The permeability of the quenched alloy begins to change at very low magnetizing forces, but that of the baked alloy, though lower, remains constant for magnetizing forces up to 3 oersteds.

Hysteresis loops for this alloy in the two conditions are shown in Fig. 10 for maximum flux densities of less than 1,000 and more than 5,000 gausses. For the baked alloy the hysteresis loops for maximum flux densities less than 1,000 gausses cannot be measured by ordinary ballistic methods, because the two sides of the loop coincide in a straight line. For loops with higher maximum flux densities the area begins to appear, but the two branches of the loop still meet at the origin. Although the coercive force is sensibly zero for the baked alloy until the maximum flux density exceeds 5,000 gausses, the hysteresis loss represented by the loop may become considerably greater than that for the quenched alloy.

7–45–25 Mo-Perminvar

The extremely low hysteresis loss and constancy of permeability at low flux densities makes 45–25 perminvar a suitable material for

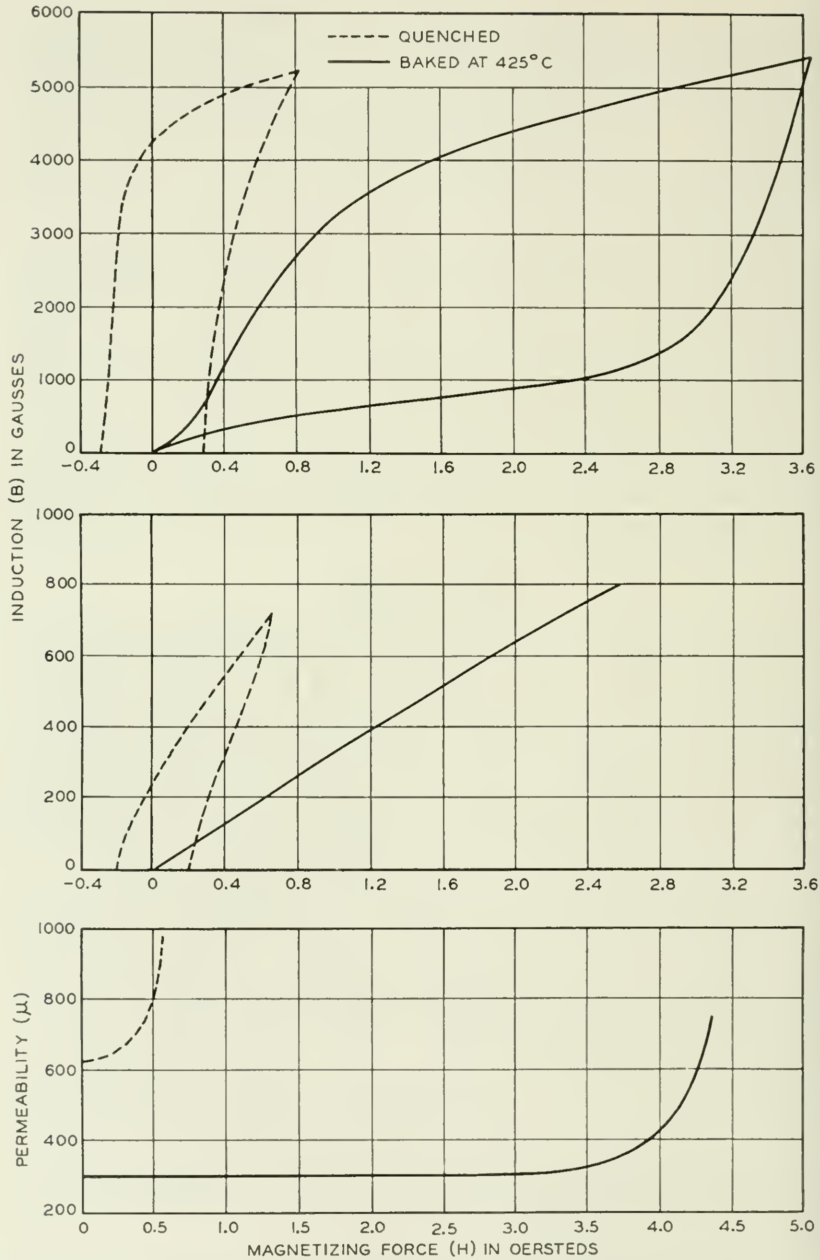


Fig. 10—Hysteresis loops and permeability curves for 45–25 perminvar.

applications where distortion and energy loss are fatal to good quality of transmission. The resistivity of this alloy is only 18 microhms-centimeter, but it can be increased without serious sacrifice of the low hysteresis characteristic by adding molybdenum. The alloy chosen for commercial use is 7–45–25 Mo-perminvar, having a resistivity of 80 microhms-centimeter.

The manner in which molybdenum affects the magnetic properties is illustrated in Figs. 3, 4, and 6. The permeability is not quite so independent of the magnetizing force as for the alloy without molybdenum, nor is the hysteresis loss quite so low. The initial permeability for the alloy baked the customary 24 hours is somewhat higher. When baked for a longer period the magnetic characteristics tend more toward those of 45–25 perminvar.

Permendur

An alloy in the iron-cobalt series used in communication apparatus is permendur. The typical composition is 50 per cent iron, 50 per cent cobalt. The outstanding magnetic property of this alloy is high permeability in the range of flux densities between 12,000 and 23,000 gausses (Figs. 4 and 6). The high permeability of this alloy "endures" to higher flux densities than does the permeability of any other magnetic material. Its initial permeability is about 700, though values as high as 1,300 have been observed for some samples. In addition to the high permeability at high flux densities permendur also has a relatively flat "butterfly" curve, as may be seen in Fig. 7. For a superposed d.-c. magnetizing force of 10 oersteds the a.-c. permeability of "Armco" iron is 40, as compared with 200 for permendur.

1.7 V-Permendur

Permendur is difficult to roll into sheets, because of its brittleness. To overcome this difficulty 1.7 per cent vanadium is added. With this addition it may be rolled into sheets as thin as a few thousandths of an inch. Th s amount of vanadium affects the magnetic properties only slightly, although larger amounts decrease the permeability at high flux densities.

Another improvement incident to the addition of vanadium is a fourfold increase of resistivity from its value of 6 microhms-centimeter for simple permendur. Permendur, it may be noted, has the lowest resistivity of the iron-cobalt series.

Laboratory Results

As stated hereinbefore, all the alloys that have been discussed have been made on a factory scale for some use in the telephone plant.

This paper, however, would be incomplete without mention of some of the remarkable magnetic properties obtained in laboratory samples. In table III some of these magnetic achievements are tabulated. The special treatments given these specimens are noted also in this table.

TABLE III

Some Remarkable Magnetic Properties Obtained in Laboratory Specimens *

Material	μ_0	μ_m	H for μ_m	H_c	Heat Treatment
"Armco" iron	20,000[1]	340,000[1]	0.021	0.03[1]	18 hr. at 1,480 deg. C., followed by 18 hr. at 880 deg. C., both in hydrogen
45 permalloy	11,000	227,000	0.025	0.0145	Melted in vacuum; electrolytic iron and electrolytic nickel; 18 hr. at 1,300 deg. C. in hydrogen
65 permalloy	2,500	610,000[2]	0.0148	0.012[2]	18 hr. at 1,400 deg. C. in hydrogen; heated to 650 deg. C. 1 hr., cooled in magnetic field of 16 oersteds in hydrogen
78.5 permalloy	13,000[3]	405,000[3]	0.0101	0.0153	$2\frac{3}{8}$ x 2 x 0.109 in. tape; annealed; wrapped in 2 layers of 3 mil. tape and quenched from 600 deg. C. in tap water
3.8–78.5 Mo-permalloy	34,000[1]	140,000[1]	0.025[1]		1,400 deg. C. in hydrogen
Permendur	1,000	37,000	0.22	0.20	940 deg. C. in hydrogen for 18 hr., slowly cooled to room temperature
Permendur	1,300	29,000	0.27		940 deg. C. in hydrogen, 6 hr.; slowly cooled to room temperature
45–25 perminvar		189,000	0.052	0.059	Heated to 1,000 deg. C. in hydrogen, reheated to 700 deg. C. and cooled in hydrogen in a magnetic field of 14 oersteds

For explanation of symbols in headings see footnote to table II.

* Except as noted, the values in this table have not previously been published.

Engineering Applications

One of the first uses of the permalloys was for continuous loading of a telegraph cable between New York and the Azores laid in 1924. For this project 78.5 permalloy was used in the form of a 0.125 by 0.006 inch tape wrapped helically on a stranded copper conductor. The average initial permeability of this alloy in the laid cable was 2,300,

considerably less than can be obtained under the best conditions of heat-treatment and absence of strains. With this loading, the speed of transmitting messages was increased fivefold.[4] By the time a second cable project was undertaken the chromium permalloys had been developed, and 2–80 Cr-permalloy was selected. This alloy has a resistivity of 45 microhms-centimeter, and the initial permeability of the loading on the laid cable was in the neighborhood of 3,700. The increase in permeability and in resistivity increased materially the message carrying capacity.[5]

The largest use of permalloys in the telephone plant has been in cores of loading coils,[6] where the alloy is used in the form of compressed insulated dust. Iron dust cores had been standard for these coils.[7] The lower magnetic losses of permalloy dust, however, permitted utilizing higher core permeabilities. This has resulted in a very material decrease in the size of loading coils. For a high grade loading coil core made from iron dust the effective core permeability at low flux densities had to be limited to 33. The first permalloy used for loading coil cores was 80 permalloy. The insulated and compressed core was designed for an effective permeability of 75—more than double that for the iron dust. Development work on an improved compressed magnetic dust core in which molybdenum is used, is now approaching completion. It is expected that the new material will have a substantially higher permeability than that of the 80-permalloy dust cores, and that it will have intrinsically superior eddy current and hysteresis loss characteristics. By virtue of these properties, it will be practicable to make a further substantial reduction in the size of loading coils without sacrifice in service standards. The decrease in the size of the cores with improvement in the core material is illustrated in Fig. 11.

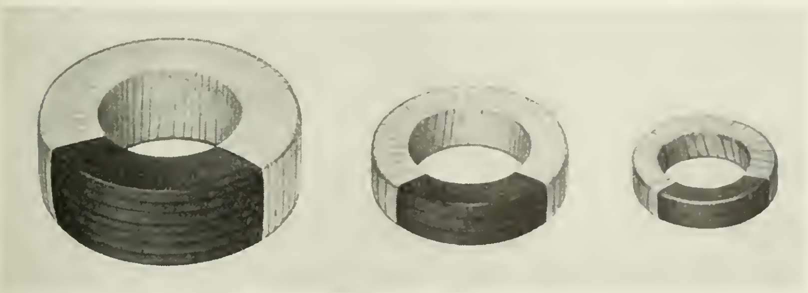


Fig. 11—Equivalent cores for loading coils. Iron dust core (left); permalloy dust core (center); molybdenum permalloy dust core (right).

[4] For all numbered references see list at end of paper.

In d.-c. apparatus where high permeability and low coercivity are of importance, and where high resistivity does not add to the usefulness of the core material, 78.5 permalloy is suitable. It is used in certain relay structures, usually of marginal type, in which the difference between operating and releasing currents is small.

For audio transformers, for retardation coils, and for other apparatus in which high permeability and high specific resistance must be combined, both 3.8–78.5 Cr-permalloy and 3.8–80 Mo-permalloy have been used. The former has slightly higher resistivity, but the latter has higher initial permeability and is more ductile.

While the initial and maximum permeabilities of 45-permalloy are not as high as those of 78.5-permalloy, the higher flux densities attained by the former and its higher resistivity favor its use for certain types of relays and transformers where high flux densities are required. It is used also in some instances for cores of coils that require high a.-c. permeability when d.-c. magnetizing forces are superposed.

The magnetic characteristics of the perminvars make them especially suitable for use in circuit elements in which distortion and energy loss must be a minimum; but their relatively high cost, and the advisability of avoiding high magnetization throughout the life of the apparatus, have prevented their extensive use in telephone plant. One use for which perminvar is especially suitable is the loading of long submarine telephone cables. Here a high resistivity is very desirable, which has been shown to be obtainable in the 7–45–25 Mo-perminvar. The increase in resistivity resulting from the addition of molybdenum more than offsets the accompanying increase in hysteresis loss, and results in a continuous loading material satisfactory for certain types of loaded cables.

Permendur was developed for use in apparatus where very high flux densities are desired. For a moderate magnetizing force flux densities of 18,000 and 23,000 gausses readily are obtained. It is used for cores and pole pieces in loud speakers, certain telephone receivers, light valves, and similar apparatus.

It may be seen from this survey that there is a great variety of magnetic materials with widely different properties from which an engineer may choose in designing magnetic elements in which magnetic flux changes are essential. Already these alloys have an important place in telephone plant. However, iron and silicon steel still are used extensively, and will continue to hold their own on a cost basis for some purposes. There is no doubt, however, that alloys of iron, nickel, and cobalt will continue to supplant iron and silicon steel in many places where circuits and apparatus are redesigned to take full advantage of their magnetic properties.

References

1. Unpublished results, P. P. Cioffi.
2. J. F. Dillinger and R. M. Bozorth. Heat Treatment of Magnetic Materials in a Magnetic Field. *Physics*, Vol. 6, September, 1935, p. 279–91.
3. G. W. Elmen. Magnetic Alloys of Iron, Nickel, and Cobalt. *Franklin Inst. Journal*, Vol. 207, 1929, p. 583–617.
4. O. E. Buckley. The Loaded Submarine Telegraph Cable. *A. I. E. E. Journal*, Vol. 44, August 1925, p. 821–9.
5. O. E. Buckley. High Speed Ocean Cable Telegraphy. *Bell System Technical Journal*, Vol. 7, 1928, p. 225–67.
6. W. J. Schackelton and I. G. Barber. Compressed Powdered Permalloy, Manufacture and Magnetic Properties. *A. I. E. E. Transactions*, Vol. 47, April, 1928, p. 429–36.
7. B. Speed and G. W. Elmen. Magnetic Properties of Compressed Powdered Iron. *A. I. E. E. Transactions*, Vol. 40,1921, p. 1321–59.

Improvements in Communication Transformers *

By A. G. GANZ and A. G. LAIRD

The rapidly advancing art of electrical communication and the increasingly wide variety of its applications have required marked improvements in the transformers used in communication circuits. These improvements, achieved partly through advances in design and partly through improvements in the constituent materials, are discussed in this paper.

THE rapid development of the art of electrical communication in the last decade has necessitated marked improvements in the transformers used in it. New applications for these transformers and the extension of old ones have imposed new and far severer performance requirements. The primary applications of communication transformers are in the telephone plant, in the various voice and carrier transmission circuits, and in a multitude of incidental services. They have also wide uses in radio broadcasting transmitters and receivers, in the amplifiers of sound motion picture equipment, in the radio equipment for aircraft, and in a variety of other circuits.

Although communication and power transformers have a common origin, the communication transformer now has evolved as a precision device which has only a general resemblance to the usual power transformer. Some voice-frequency transformers, such as those used in aircraft, weigh but 2 or 3 ounces, yet transmit speech substantially undistorted. Some used in program circuits transmit with negligibly small phase or amplitude distortion all frequencies from 20 to 16,000 cycles per second. Transformers also have been developed for transmitting narrow bands of frequencies and having associated with the normal transformer performance valuable frequency discriminating properties. A discussion of improvements in these narrow band transformers is outside the scope of this paper, which will be confined to those transmitting wide-frequency bands, that is, those for which the ratio of upper to lower limiting frequencies is at least 10 to 1.

The design of the modern communication transformer is based upon extensions of the familiar theory of transformers covered in numerous texts. However, this type of transformer is a more complex device, with its multiplicity of requirements and its transmission over a wide-frequency range. Its proper representation accordingly requires a

* Published in the December 1935 issue of *Electrical Engineering*, and scheduled for presentation at the Winter Convention of the A. I. E. E., New York City, January 28–31, 1936.

more elaborate equivalent network than the customary Π or T network. With the use of such a network, the performance of the transformer may be correlated mathematically with its constants in accordance with network theory.

Improvements in Low-Frequency Transmission

The great improvements in communication transformers, particularly at audio frequencies, are largely attributable to the invention and application of the permalloys as magnetic core materials. For convenience, the term "permalloy" has been applied to a group of nickel-iron alloys containing between 30 and 95 per cent nickel which have been developed by Bell Telephone Laboratories.[1, 2, 3] In addition to other desirable magnetic properties, some of these permalloys [4] when properly heat treated yield exceptionally high initial permeabilities. As a result of the use of these special alloys, telephone transformers may be designed to have less loss and distortion over wider frequency ranges than has been possible in transformers designed without the benefit of use of such materials.

Figure 1 illustrates the excellence of performance resulting from the use of a special permalloy consisting of approximately 4 per cent

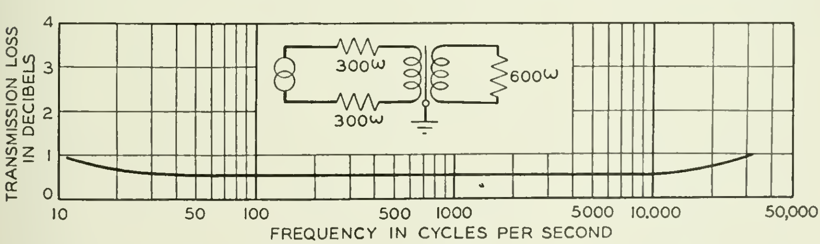


Fig. 1—Transmission-frequency characteristic of a transformer utilizing a permalloy core and designed to connect a telephone transmission line to a program repeater. Transmission loss shown is the loss relative to an ideal transformer of the same ratio.

chromium, 78 per cent nickel, and the remainder iron, in a transformer designed to connect a telephone transmission line to a program repeater. Figure 2 shows the voltage amplification characteristic of an interstage transformer for a high quality amplifier. As is well known,[5] superimposed direct currents generally decrease the effective a.-c. permeability of ferromagnetic materials. Therefore, to retain the full benefit of the permalloy core, an auxiliary circuit was used with this transformer for supplying the plate current to the preceding tube. Another illustration is a transformer designed for use as an input transformer (that is, one designed to operate into the grid circuit of a

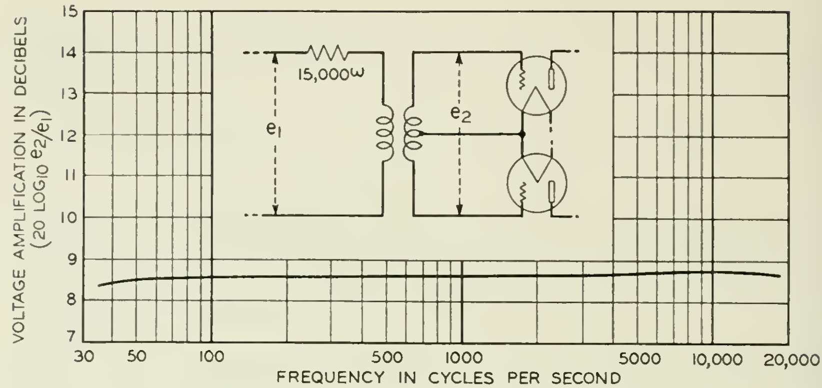


Fig. 2—Voltage amplification-frequency characteristic of an interstage transformer for a high quality program amplifier.

vacuum tube) for portable recording equipment where light weight is an important consideration. The voltage amplification-frequency characteristic of this transformer is shown in Fig. 3. There is shown also in this figure, for purposes of comparison, the very much poorer characteristic realized when an alloy of about 50 per cent nickel as commonly used is substituted for the special permalloy. In addition,

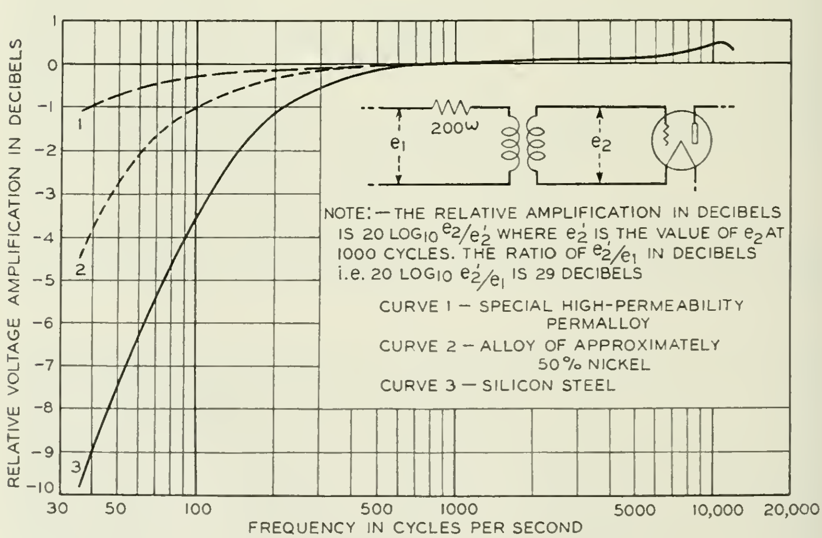


Fig. 3—Curves showing the relative effect of different core materials on the voltage amplification-frequency characteristic of a small transformer designed for portable recording apparatus. The composition of the special permalloy is approximately 4 per cent chromium, 78 per cent nickel, and the remainder iron.

there is shown the effect of using silicon steel as core material, which was the practice in older transformers.

An important limitation of high permeability alloys is their sensitivity to mechanical strain which may seriously impair their magnetic characteristics. Considerable care must be exercised to avoid strain during assembly operations after laminations are annealed. Telephone transformers are designed specially to provide a firm assembly without mechanical strain, thereby retaining the high permeabilities available.

Increase in Voltage Amplification

As may be seen from the foregoing curves, the voltage amplification of input transformers at the low end of the frequency band is directly dependent on the permeability of the magnetic core material. At the highest frequencies the voltage amplification of the above input transformers is controlled by leakage and capacitances, the latter including grid circuit capacitances as well as the transformer distributed capacitances. Over a wide range in the central part of the frequency band these effects are negligible and the transformer performs much as an ideal transformer of the same ratio. By proper proportioning of the leakage and capacitance effects, the shape of the characteristic may be controlled to a certain measure at will. For example, a rising voltage amplification-frequency characteristic can be obtained if desired to correct for a falling characteristic of other parts of the amplifier.

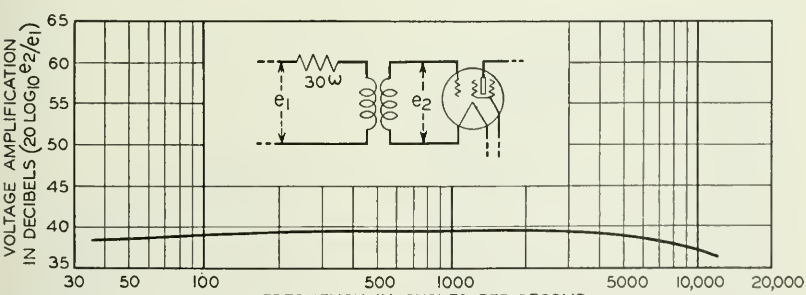


Fig. 4—Voltage amplification-frequency characteristic of an input transformer having an impedance ratio of 10,000 to 1.

In certain types of circuits the voltage amplification of input transformers is at a high premium, such as in the amplification of low-energy signals when a.-c. power is used for the tubes. Under these conditions the tubes tend to introduce appreciable noise. A high-voltage amplification in the input transformer serves to raise the signal voltage at the grid terminals so as to override the tube noises. Figure 4 shows

that a high amplification, in this instance 10,000 to 1 in impedance level from a 30-ohm source, may be realized without undue restriction in either the low- or the high-frequency transmission.

Since transformers of this type are located at points of very low energy levels, special pains must be taken to avoid interference from stray magnetic and electrostatic fields. To prevent hum from nearby power apparatus, transformers are enclosed in cases or shields of high permeability material. The interference voltages induced in these transformers are some 30 or 40 decibels less than in older unshielded types of transformers. For higher frequency interference, effective shielding is obtained by cases made of high-conductivity material such as copper or aluminum.

Reduction in Size and Weight

The demand for lightweight equipment for aircraft communication and for portable apparatus for testing and recording has resulted in the development of communication transformers of unusually small size and weight. The smaller sizes weigh only $3\frac{1}{2}$ ounces and occupy a space of but 3 cubic inches. One of these used in aircraft receiving sets is illustrated in Fig. 5 contrasted with an earlier transformer also for lightweight service. The transmission loss-frequency characteristics are shown in Fig. 6. The corresponding characteristic of an input transformer for similar service is shown in Fig. 7.

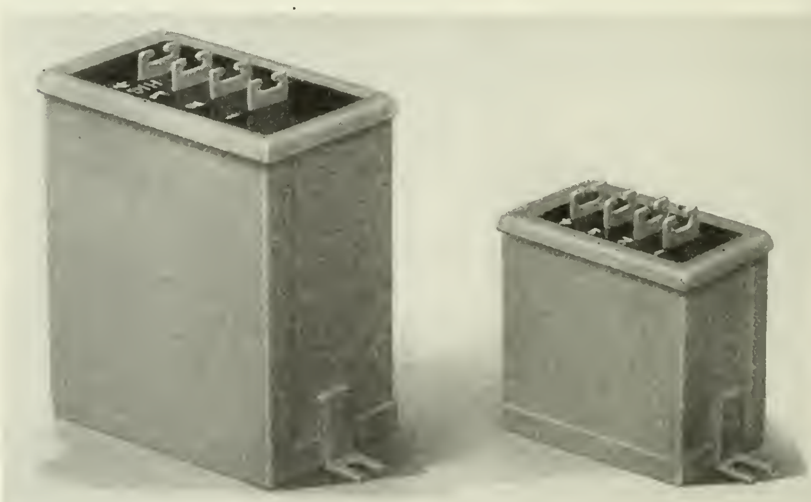


Fig. 5—Output transformer A (right) utilizing a permalloy core transmits frequencies from 40 to 3,000 cycles with greater over-all efficiency than the larger output transformer B (left) utilizing a core of silicon steel (see Fig. 6).

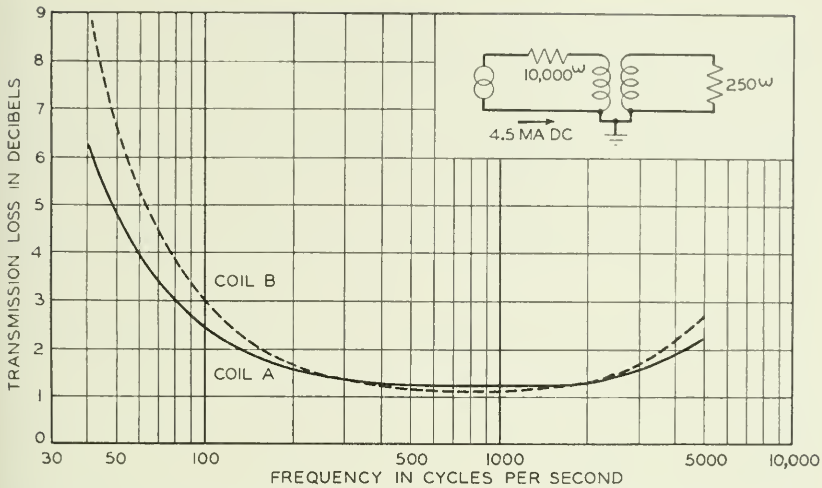


Fig. 6—Curves showing the transmission loss-frequency characteristics of (*A*) the small output transformer and (*B*) the larger output transformer shown in Fig. 5.

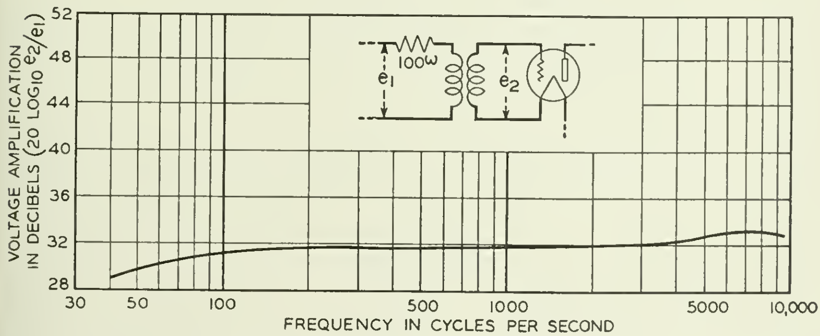


Fig. 7—Voltage amplification-frequency characteristic of an input transformer similar in size to the smaller output transformer shown in Fig. 5.

Extension of Range to Higher Frequencies.

The effective permeabilities of alloys of high initial permeabilities drop rather rapidly with frequency, a property which lessens the value of such alloys in transformers for carrier and higher frequencies. This effect, which is attributable primarily to eddy currents, can be greatly decreased, of course, by the use of thinner laminations. The use of these alloys, however, is limited by the rapidly increasing cost of reducing the lamination thickness and the less efficient use of the volume available for the core. This dropping off of effective permeability with frequency is not so important in audio-frequency trans-

formers, since there the core characteristics limit the transmission only at low frequencies where the effective permeability is high.

At higher frequencies the voltage amplification is severely limited also by the grid circuit and transformer capacitances. It has been found advantageous to add correcting elements, such as inductances and capacitances, to increase the gain ordinarily available. These added elements and the equivalent elements in the transformer are designed together as configurations similar to low-pass filter sections terminated midshunt in the grid circuit capacitance. Figure 8 illustrates the performance of a transformer designed for certain high frequency transmission experiments. The performance of this transformer in the frequency range of 30–500,000 cycles per second is similar to that of earlier types in the range of 30–60,000 cycles per second.

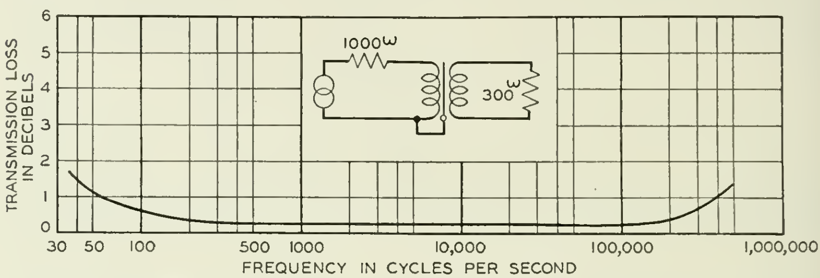


Fig. 8—Transmission loss-frequency characteristic of a transformer designed to transmit high frequencies in addition to audio frequencies.

In most telephone applications the design of input transformers is complicated greatly by circuit functional requirements in addition to those of direct transmission. For example, as discussed in a succeeding section of this paper, phase distortion requirements demand self-impedances far higher than those required by transmission loss considerations, thermal [6] noise requirements demand lower dissipations, and impedance requirements limit the voltage amplification obtainable. For feed-back type amplifiers [7] the input transformers in addition to the forward amplification must meet transmission and phase requirements in the regenerative path. These special requirements apply not only for the transmitted frequency band but also at frequencies remote from this band in order to insure stability of the amplifier.

Reduction in Phase Distortion

Since transformers are reactive devices, they introduce phase shift in the circuits in which they are used. If the phase shift introduced be a linear function of the frequency it will not produce any distortion

in the shape of the transmitted wave. However, departures from linearity change the wave shape, and this form of distortion is referred to as phase or delay distortion. The delay at any frequency is a measure of this departure from linearity, and is dependent upon the frequency derivative of the phase shift at that frequency. Differences in delay of the various frequency components of the signal wave which transformers tend to produce result in distortion that may be especially serious in circuits intended for program transmission.

For wide-band transformers the delay caused by the shunting effect of the mutual impedance usually predominates. In fact for audio transformers the delay at higher frequencies is relatively so small that the delay distortion is practically equal to the mutual impedance delay at the lowest transmitted frequency. Delay distortion is also of importance in transformers to be used in television and telephotography. In these circuits phase distortion causes a space shift in the image of certain frequency components with respect to others with consequent blurring of the image.

The delay characteristic of a transformer used in program circuits to connect a telephone transmission line to the grid circuit of a repeater amplifier is shown in Fig. 9. This characteristic is compared in the same illustration with the delay characteristic of a repeater transformer developed some years ago for use in what then was re-

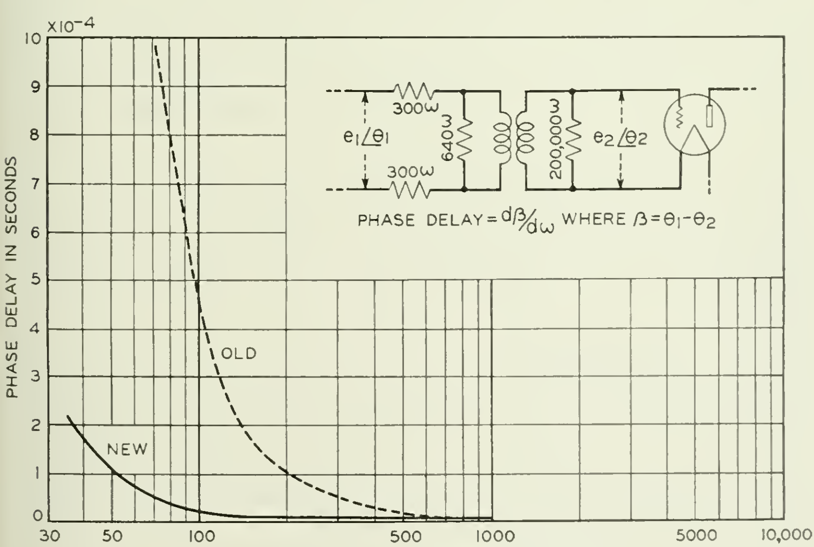


Fig. 9—Phase delay-frequency characteristic of an input transformer used in recent program repeaters, compared with that of an input transformer used in an older repeater.

garded as a high quality circuit. The delay characteristic of a high frequency transformer is shown in Fig. 10.

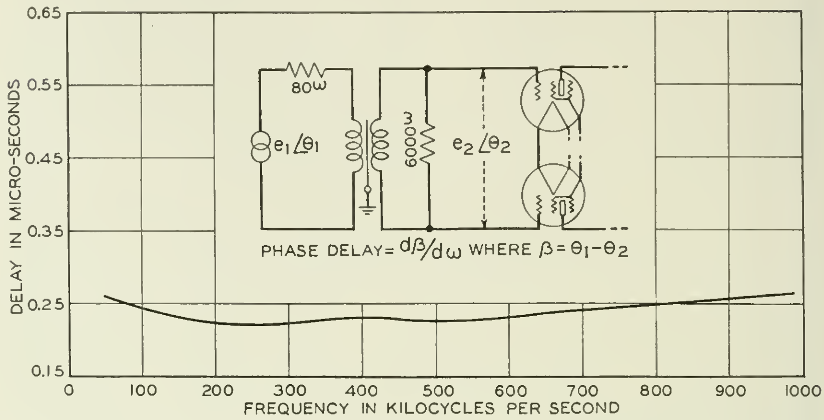


Fig. 10—Phase delay-frequency characteristic of an input transformer designed to transmit radio frequencies.

Reduction in Audio Frequency Modulation

The present exacting requirements for transformer performance have made it necessary to lessen greatly certain second-order distortion effects inherent in transformers having magnetic cores. Nearly all core materials tend to generate extraneous frequencies because of magnetic non-linearity—a property referred to as magnetic modulation. In audio-frequency circuits intended for high quality service, magnetic modulation may cause serious distortion in that harmonics of the lower frequencies appear higher in the audible range. The energy present at the lower frequencies is usually so much greater than that over the rest of the band that the modulation products may approach the order of magnitude of the signal components at higher frequencies.

This form of distortion in no way is revealed by the ordinary transmission loss characteristic; in fact, a transformer having a very flat loss characteristic over a wide-frequency range may nevertheless be definitely objectionable from a modulation standpoint. In present audio-frequency transformers, the total modulation products are some 40 to 80 decibels down from the energy of frequencies around 35 cycles per second producing them. This represents an improvement of about 30 decibels over older types.

Another second-order effect resembling modulation is microphonic noise caused by magnetostriction phenomena, that is, changes in magnetization accompanying the physical deformation of the magnetic

material. For instance, slight jarring of input transformers used in very high-gain amplifiers (100 decibels or more) may induce in this manner disturbing noise voltages. Freedom from magnetostriction is a unique characteristic of permalloys containing approximately 80 per cent nickel, and their use accounts for the superiority of telephone transformers from this standpoint.

Reduction in Carrier Frequency Modulation

Magnetic modulation is even more serious at carrier frequencies where a transformer may transmit many channels, frequently at widely different levels. The modulation from the higher level channels may produce very objectionable interference in other channels. Carrier transformers now have been improved to such an extent that highly sensitive testing circuits are required to detect and measure the modulation products contributed by them. Representative values of these modulation products expressed as current ratios to the fundamentals are of the order of one-millionth of the fundamental frequencies, compared to one-thousandth in older types.

It is of interest to point out that in such transformers the presence of magnetic material, other than the special permalloys, in the vicinity of the transformer must be avoided with great care. For example, a small steel screw near the field of the transformer will seriously impair its performance from a modulation standpoint. Common practice is to use brass parts for the assembly and to confine the field of the transformer by completely enclosing it in a copper or aluminum case. The transformer then may be mounted by any convenient means without affecting its performance.

Improvements in Shielding and Balance

In the very nature of the service that it renders, the telephone plant involves many independent communication circuits in fairly close proximity. The minimizing of interference between these independent communication circuits has constituted a major problem in telephone engineering. Where such interference occurs between like circuits, that is, between two voice circuits or between two carrier circuits of overlapping frequency bands, the interference commonly is referred to as crosstalk, as distinguished from the interference from other types of circuits such as power and telegraph circuits.

In order to avoid crosstalk and other interference, balanced * cir-

* The new coaxial cable circuits under development are an interesting exception. In this type of cable a grounded outer conductor completely encloses the central conductor, and shielding rather than balance is relied upon to protect the circuit from interference. The shielding depends upon the size, thickness, permeability, and conductivity of the outer conductor and the frequency of the disturbance. If the frequency band used in the transmission over coaxial circuits is chosen properly, the circuit may be made substantially immune to effects from outside fields.

cuits are used almost exclusively in the telephone plant. For simplicity, some of the terminal apparatus connected with these circuits is constructed unbalanced, so that it is necessary to interpose transformers between the lines and the office equipment, these transformers providing a barrier to the propagation of the relatively large longitudinal currents from the line circuits. (Longitudinal currents, in contrast with the usual circulating currents, are currents that flow equally and in the same direction in both sides of the line.) In order to insure that the voltage impressed on the office equipment is attributable to the voltage between the wires of the line circuit and not to that between the wires and ground, it is necessary that the transformers be balanced very carefully; and for certain types of circuits, shields must be interposed so that the direct capacitance between the line winding and office winding is reduced to a very small value.

With a greater emphasis on carrier frequency transmission, a higher degree of balance is required between certain transformer windings, and highly effective shielding is frequently necessary. It is necessary also that the line windings be balanced very closely with respect to capacitances to the shield and case. The unbalance effects in carrier transformers now have been reduced to values in the order of 1 or 2 microamperes in circulating current per volt between the line windings and ground at 30,000 cycles per second, which compares with values of 50 microamperes or more for older transformers. At the same time the electrostatic shielding between the windings has been improved to such an extent that the direct capacitance between windings has been reduced to 1 or 2 micromicrofarads instead of 30 or 40 as before. The shields are arranged to intercept the dielectric flux lines tending to connect the primary and secondary windings, so that the direct capacitance between the two windings is attributable only to stray flux which bypasses the shield. One of the windings usually is enclosed completely in lead or copper foil with overlapping edges insulated to prevent a short-circuited turn. Still further improvements are obtained by covering the leads with grounded metal braiding, and in special cases by enclosing the terminals of the shielded winding in a separate shielded compartment. In certain transformers designed for high precision testing equipment, the direct capacitance between windings has been reduced to values less than 0.001 micromicrofarad.

In connection with phantom circuits, severer crosstalk requirements have necessitated more precise balances in the associated voice frequency transformers. In these transformers the turns are so arranged that the various distributed capacitances, flux linkages, and d.-c. resistances are disposed symmetrically with respect to ground. It has

been found in practice that this symmetry is realized most readily by close coupling between the various parts of the windings. By improvements in design, the crosstalk between phantom and side circuits has been reduced to values in the order of 20 millionths in current ratio, compared to values 5 times as large, formerly tolerated.

Reduction in Impedance Distortion

As a further consequence of the extension of carrier systems, it has become necessary to match the impedance presented by transformers, when terminated in the succeeding circuits, to particular values over the frequency range. For example, the transposition schemes used on open wire lines are such as to minimize crosstalk primarily for carrier signals propagated in one direction in any line. If the transformer terminating such a line does not present an impedance under load equal to the characteristic impedance of the line, a portion of the wave is propagated in the reverse direction, that is reflected, causing crosstalk into adjacent circuits. This reflection effect increases with the vector difference in the impedances of the transformer and the line, the latter impedance approaching a pure resistance as the frequency increases.

The impedance of transformers has become increasingly important where such transformers terminate filters that require a nearly constant resistance termination to maintain proper attenuation characteristics. Another example is in transformers terminating screen grid tubes where the plate impedances are relatively very high. Here the energy abstracted from the plate circuit and transmitted by the transformer is directly dependent upon the resistance component of the impedance of the transformer when terminated in its load.

Better impedance characteristics of transformers for these various applications have been obtained by increasing the mutual impedance and decreasing the leakage and capacitance effects. This procedure is made difficult by the necessity for meeting at the same time other and newer requirements, as, for example, modulation limits. Correcting elements consisting of capacitances and inductances usually are added and are proportioned with the transformer elements in accordance with network theory. Typical impedance characteristics of such transformers are shown in Figs. 11 and 12.

Input transformers operating into the grid circuits of vacuum tubes inherently have impedances that depart widely from the nearly pure resistances usually desired, because of the reactive termination provided by the grid circuit. This makes it necessary to add resistances to serve in place of the usual load resistance. The required dissipation

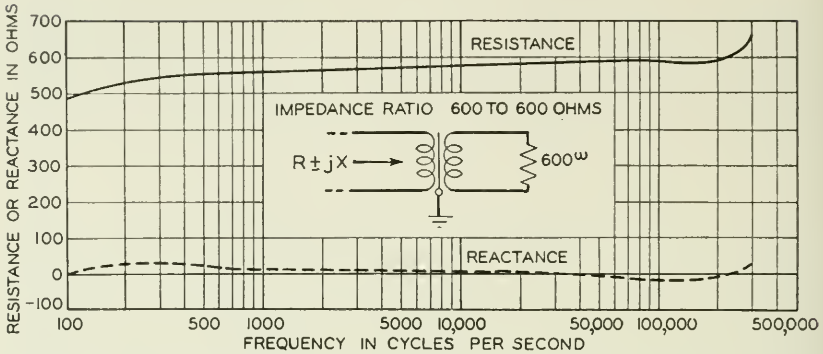


Fig. 11—Impedance-frequency characteristic of a transformer designed to operate between 600-ohm terminations and to transmit high frequencies in addition to audio frequencies.

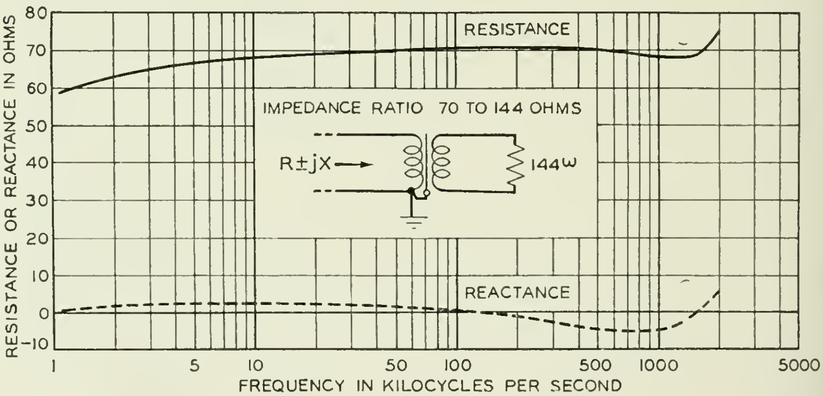


Fig. 12—Impedance-frequency characteristic of a transformer designed to operate between resistance terminations of 70 ohms and 144 ohms.

may be provided in many ways in the input transformer, which consideration allows much wider latitude in the design than in the ordinary transformer; there the major part of the dissipation for low transmission loss necessarily must occur in the load. A typical impedance characteristic of such an input transformer is shown in Fig. 13.

Testing Precautions

As a necessary concomitant to improvements in transformers, more precise testing circuits have been developed for accurately determining transformer performance. In estimating the performance of the transformer from its characteristic, care must be taken to make sure that the service conditions were reproduced carefully in the measuring circuit. In particular, certain precautions must be observed in the

measurement in order to avoid obtaining a misleading characteristic. For example, the permeability of magnetic core materials tends to rise rapidly from its initial value with increasing voltages. If in the measurement, the low-frequency voltages used are materially higher than they are under service conditions, the low-frequency response will appear to be much better than the true response.

As another example, the transmission of input transformers at the high-frequency end may be critical with the termination of the high-

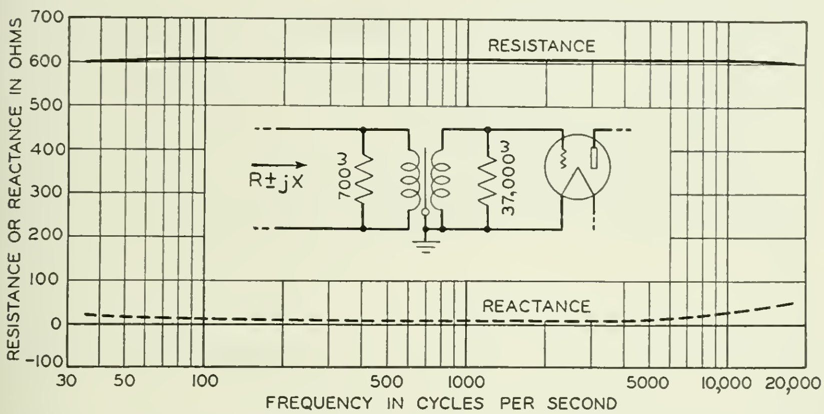


Fig. 13—Impedance-frequency characteristic of an input transformer used to operate from a telephone transmission line into a high quality program repeater. Impedance ratio is 600 to 15,000 ohms.

voltage winding. If the grid capacitance and conductance conditions are not reproduced faithfully in the measuring circuit, the high-frequency voltage amplification may appear to be much better than the true value.

In addition to more precise transmission measuring circuits, various other special circuits have been developed for measuring transformers, such as modulation, impedance, and crosstalk measuring circuits. The design of these circuits is of necessity a specialized art.

In the foregoing, various types of improvements in communication transformers have been discussed. Wherever applicable, several such improvements have been incorporated in an individual design. The improved performance of transformers as described has been an essential step in the development of the communication art.

References

1. H. D. Arnold and G. W. Elmen. Permalloy—an Alloy of Remarkable Magnetic Properties. *Franklin Inst. Journal*, May 1923, p. 621–32.
2. G. W. Elmen. Magnetic Alloys of Iron, Nickel, and Cobalt. *Franklin Inst. Journal*, May 1929, Vol. 207, p. 595.

3. G. W. Elmen, U. S. Patents: 1,586,883; 1,586,884; 1,620,878; etc.
4. G. W. Elmen, U. S. Patent, 1,768,237.
5. C. R. Hanna. Design of Reactances and Transformers Which Carry Direct Current. *A. I. E. E. Journal*, Vol. 46, February 1927, p. 128–31.
6. J. B. Johnson. Thermal Agitation of Electricity in Conductors. *Phys. Rev.*, July 1928, p. 97–109.
7. H. S. Black. Stabilized Feedback Amplifiers. *Elec. Engg.* (*A. I. E. E. Trans.*), Vol. 53, January 1934, p. 114–20.
8. W. L. Casper. Telephone Transformers. *A. I. E. E. Journal*, Vol. 43, March 1924, p. 197–209.
9. T. Spooner. Audio Frequency Transformer Characteristics. *Elec. Journal*, July 1926, p. 367–75.
10. F. E. Field. The Evolution of the Input Transformer. *Bell Lab. Record*, Vol. 3, October 1926.
11. E. L. Schwartz. Permalloy in Audio Transformers. *Bell Lab. Record*, Vol. 6, April 1928.
12. K. S. Johnson. Transmission Circuits for Telephonic Communication (a book). D. Van Nostrand Company, Inc., 1927.
13. T. E. Shea. Transmission Networks and Wave Filters (a book). D. Van Nostrand Company, Inc., 1929.
14. C. E. Lane. Phase Distortion in Telephone Apparatus. *Bell Sys. Tech. Jour.*, v. 9, July 1930, p. 493–521.
15. Eugene Peterson. Harmonic Production in Ferromagnetic Materials. *Bell Sys. Tech. Jour.*, v. 7, Oct. 1928, p. 762–96.

Measurement of Telephone Noise and Power Wave Shape *

By J. M. BARSTOW, P. W. BLYE and H. E. KENT

IN studies of the inductive coordination of power and telephone systems from the noise standpoint, a knowledge of the magnitudes of the harmonic currents and voltages on the power circuits and of the harmonic components of the telephone circuit noise is necessary. It is also necessary that there be available a means of rating and summing up these individual components to give an overall indication of their effects on a person using a telephone connected to one of the exposed circuits. This paper discusses methods which have been developed for making such overall measurements.

The effect on a listener of a given amount of noise on a telephone circuit is a complex one, and it is not practicable in the day-by-day maintenance of telephone circuits to measure separately all the factors involved. Rather, it is necessary to make some overall measurement of the circuit noise which may be related to its effect on telephone transmission. It is, of course, desirable that the measuring devices used should measure different circuit noises as equal when they produce equal interfering effects on telephone transmission.

Two methods of measuring telephone circuit noise are at present in use in the Bell System. One of these methods is subjective, that is, uses the human hearing mechanism as a part of the measuring apparatus. This method consists of comparing, in a telephone receiver, the noise to be measured with a noise generated by means of a standard buzzer. The observer adjusts the magnitude of the buzzer noise by means of a calibrated potentiometer until, in his judgment, it is as disturbing as the noise to be measured.

The objective method of noise measurement which has been made available within the last few years employs an electrical network for weighting the various single frequency components of a noise as closely as practicable in accordance with their interfering effects on telephone transmission, and a calibrated amplifier to raise the energy level of the weighted components sufficiently to operate an electric meter. The chief operating advantages of the objective method are the repro-

* Digest of a paper published in the December 1935 issue of *Electrical Engineering* and scheduled for presentation at the A.I.E.E. Winter Convention, New York, N. Y., January 28–31, 1936.

ducibility of the results and the ease and speed of making measurements. Its disadvantage lies in the difficulty in determining the complex nature of the human hearing mechanism and simulating its characteristics sufficiently well in objective apparatus.

One of the important steps in the development of the objective noise meters has been the determination of the relative interfering effects of different single frequency tones. Two types of tests have been used for this purpose, (*a*) judgment tests and (*b*) articulation tests.

Judgment tests usually are set up so that the observer may compare directly two noises in the presence of speech heard over a representative telephone circuit. The magnitude of one of the noises is adjusted until it is judged to be as disturbing as the second noise. The magnitudes which the observer judges to be equally disturbing can be measured and in the case of single frequency tones, the relative weighting which should be applied to the two frequencies may thus be determined.

An articulation test consists essentially in calling a number of meaningless monosyllables over a circuit to a group of observers, each of whom records the sounds that he hears. The percentage of sounds correctly received is termed the "per cent articulation" for the particular condition tested. On a given circuit, two different noises which produce the same loss in articulation would usually be considered as equally interfering. As before, in the case of two different single frequency noises, measurements may be used to determine the relative weightings to be applied to the two frequencies.

In 1919, the results of judgment and articulation tests on the relative interfering effects of different single frequency tones were published in a paper by H. S. Osborne.[1] Since that time, several other sets of tests of this character have been made in order to check the values previously obtained and to extend the frequency range covered. From the results of all these tests [2] and a recognition of the trend toward more uniform frequency response in telephone message channels, a single curve of relative interfering effects of different single-frequency tones in a telephone receiver has been derived. This is shown in Fig. 1, the curve being labeled " receiver currents." By combining with this curve the frequency characteristic of a representative transmission path between the toll circuit terminals and the

[1] "Review of Work of the Subcommittee on Wave Shape Standard of the Standards Committee," H. S. Osborne, *A. I. E. E. Transactions*, Vol. 38, Part 1, 1919.

[2] The articulation and judgment tests mentioned here also contributed largely to the selection, by the C. C. I. F. (international advisory committee on telephony), of a curve of relative interfering effects of single-frequency tones expressed in terms of voltage across the receiver, which it has recommended as a basis for noise measurement on international circuits. The weighting given in curve A of Fig. 1, when expressed in similar terms is in conformity with the weighting recommended by the C. C. I. F.

telephone subscriber's receiver, a second curve indicating the relative interfering effects of different single-frequency currents in the telephone line has been derived. This is also shown in Fig. 1. These two curves have been incorporated in the indicating noise meters for use in measuring noise in the subscriber's receiver and noise at the terminals of a toll circuit, respectively.

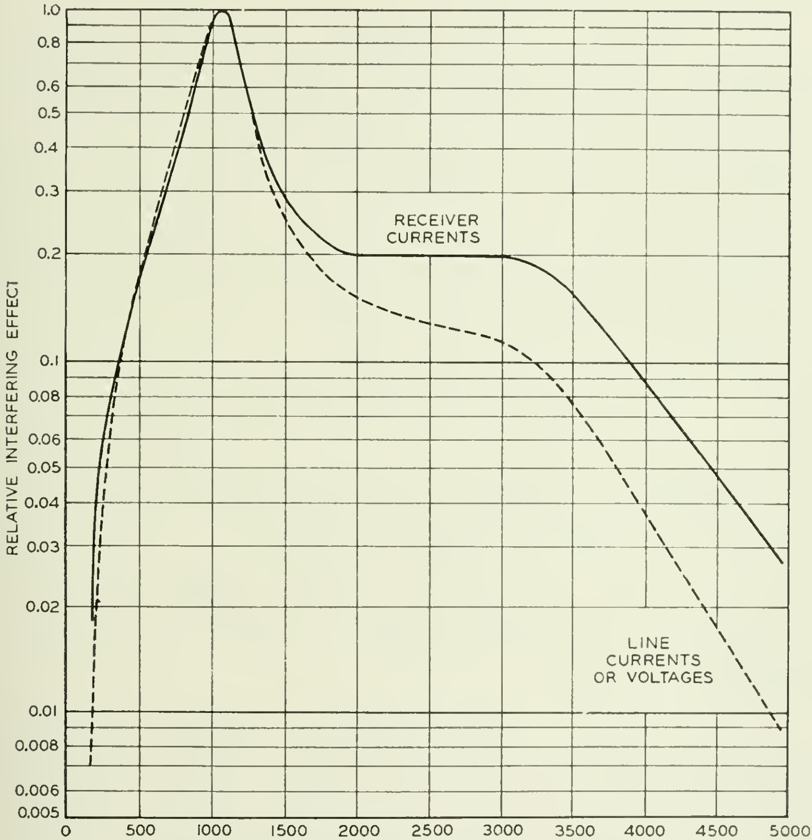


Fig. 1—Relative interfering effects of telephone circuit noise currents.

A second important factor considered was the manner in which various single-frequency noises combine in the human ear. The rule of combination adopted was that by which each single frequency contributed to the total meter reading in proportion to its weighted power. (This is the equivalent of the familiar root-sum-square rule for summing up currents or voltages,)

In addition to the requirements for weighting and rule of combination, it was thought desirable to employ an indicating instrument in which the change of reading was about as rapid as the change in appreciation of loudness in human hearing. From published results and confirming tests it was determined that on the average, the indicating instrument should reach a full deflection for sounds lasting .2 second or longer.

Under these general specifications, several models of circuit noise meters were built and two series of tests were made to determine their adequacy for measuring circuit noise. These tests were made under the auspices of the Joint Subcommittee on Development and Research of the Edison Electric Institute and the Bell System. The first was a rather extensive series of articulation tests on open-wire toll circuit noise. Since none of the toll circuit noises tested contained components of importance above 2,000 cycles, a series of judgment tests was carried out on representative noise of the type arising by induction in telephone circuits exposed to a-c. lines supplying rectifiers and on various high-frequency noises derived therefrom.

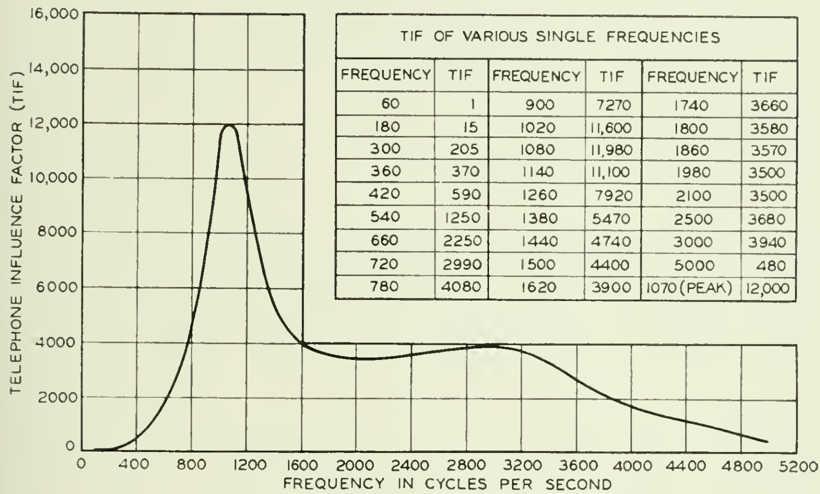
The articulation tests showed that when toll circuit noises of various types produced equal losses in articulation under the given set of telephone conditions, they were measured as substantially equal by both the objective and subjective methods of measuring. The objective method gave a slightly better correlation than did the subjective method even though the average of 18 individual observers was used in the latter. While the correlations were not as close with the high-frequency noises as in the case of the more common types of toll circuit noise, on the average the noise meter rated the rectifier noises at least as well as did the ear balance method, the latter using 10 observers.

A device called the "Telephone Interference Factor Meter" for measuring or rating the wave shape of power system currents and voltages in terms of their influence on exposed telephone circuits was described in the Osborne paper of 1919, referred to above. With this instrument, an indication was obtained of the total harmonic content of a given voltage wave, the individual components present being weighted approximately in proportion to their relative interfering effects.

The data obtained from the more recent studies of relative interfering effects described above have made possible a revision of the method of measuring T.I.F., in which the basic principle has been retained but in which the frequency weighting characteristic has been revised somewhat and its range extended to about 5,000 cycles. In

connection with this revision, the name has been changed to "Telephone Influence Factor."

The Telephone Influence Factor (T.I.F.) of a voltage or current wave is the ratio of the square root of the sum of the squares of the weighted effective values of all the sine wave components (including, in alternating current waves, both fundamental and harmonics) to the effective value of the wave. The weightings decided upon to be applied to the individual components are as shown in Fig. 2.



TIF OF VARIOUS SINGLE FREQUENCIES

FREQUENCY	TIF	FREQUENCY	TIF	FREQUENCY	TIF
60	1	900	7270	1740	3660
180	15	1020	11,600	1800	3580
300	205	1080	11,980	1860	3570
360	370	1140	11,100	1980	3500
420	590	1260	7920	2100	3500
540	1250	1380	5470	2500	3680
660	2250	1440	4740	3000	3940
720	2990	1500	4400	5000	480
780	4080	1620	3900	1070 (PEAK)	12,000

Fig. 2—Frequency weighting characteristic for TIF measurements.

In deriving the revised frequency weighting characteristic, the following factors representing distortion occurring in the various media intervening between the power circuit current or voltage and the telephone subscriber's ear were considered.

1. Relative interfering effects of single-frequency components in the receiver of a subscriber's telephone set.
2. Distortion occurring between the terminals of the circuit in which the noise is induced and the subscriber's receiver.
3. Variation in coupling between power and telephone circuits with frequency.
4. Variation of effects of telephone circuit unbalances with frequency.

Data on Items 1 and 2 above were combined to derive the line weighting characteristic of the telephone circuit noise meters indicated by the "line currents" curve of Fig. 1. It was, therefore, possible to use this curve directly to represent the combination of these two

factors. A factor directly proportional to frequency was adopted to represent inductive coupling between power and telephone circuits (Item 3), the work of the Joint Subcommittee on Development and Research having indicated that, in general, coupling may be so represented. After studying data available on Item 4, it was concluded that no type of frequency weighting could be adopted which would satisfactorily represent all types of telephone circuit unbalances. Thus T.I.F. as measured by the method described here is a correct index to the influence of a power circuit voltage or current only for those cases where unbalances are independent of frequency. This is usually the case on open-wire toll circuits and open-wire exchange circuits employing bridged ringers. In other cases some empirical modification may be necessary.

Since a large amount of data has been obtained with the old T.I.F. meter, it was considered desirable, if practicable, to adjust the scale of the revised set so that readings made by this method would, in general, be approximately the same as readings obtained by the old meter. In this connection calculations using the old and new weightings were made on a large number of machines and circuits of various types on which harmonic analyses were available. These calculations were supplemented by a considerable number of comparative measurements in the factory and in the field, using meters employing the old and new weightings. These calculations and tests indicated that in the average case, reasonably satisfactory correlation between the readings made by the two meters would result if a peak value of 12,000 were assigned to the new weighting characteristic, as shown in Fig. 2.

Several experimental models of T.I.F. measuring sets were made employing the new weighting characteristic and these have given very satisfactory results. The adoption of the rule that coupling is to be considered proportional to frequency also makes it possible to use a circuit noise meter and a small amount of auxiliary apparatus to form a T.I.F. meter.

The development of the revised method of measuring T.I.F. has also been conducted under the auspices of the Joint Subcommittee on Development and Research of the Edison Electric Institute and the Bell Telephone System.

On the Correlation of Radio Transmission with Solar Phenomena *

By A. M. SKELLETT

A DAILY character figure for radio transmission is obtained from the data of the short-wave transatlantic telephone circuits of the American Telephone and Telegraph Company. The New York-London circuits are in practically continual use so that they furnish data from which a character figure, representative of the whole 24 hours, may be derived. Such figures are based on the ratio of uncommercial to total time and thus are indirectly dependent on field strengths.

In order to facilitate plotting, these character figures were reduced to 3 group indices. Figure 1 shows the indices arranged to bring out the twenty-seven-day recurrence tendency. This is demonstrated by the apparent bunching of the spots into more or less vertical columns. Terrestrial magnetic data are shown alongside in similar form for comparison.

The recurrence tendency is well enough marked in the chart so that useful predictions of future behavior may be made. The chart is kept up to date and then by inspection a prediction may be made for any day not more than twenty-seven days distant. Some idea of the probable accuracy may also be obtained from the chart by noting whether the day in question falls, for instance, in the middle of a major sequence or on the ragged edge of a poorly defined one. Such probable accuracy is expressed by modification of the prediction with the words "probably" or "possibly."

The correlation between the two phenomena is good enough so that predictions of activity of one nature may be made from the chart of the other type of activity. For instance it would be possible to predict the radio behavior from the magnetic chart alone. This method has been found to yield the same order of accuracy as that using the radio chart alone.

Daily predictions of the behavior of the radio circuits from either the radio or magnetic chart have been correct 62 per cent of the time. Similar predictions of the magnetic data from the magnetic chart have

* Digest of a paper presented at the Sixteenth Annual Meeting of the American Geophysical Union, Washington, D. C., April 25, 1935, and published in full in the *Proceedings of the Institute of Radio Engineers*, November, 1935.

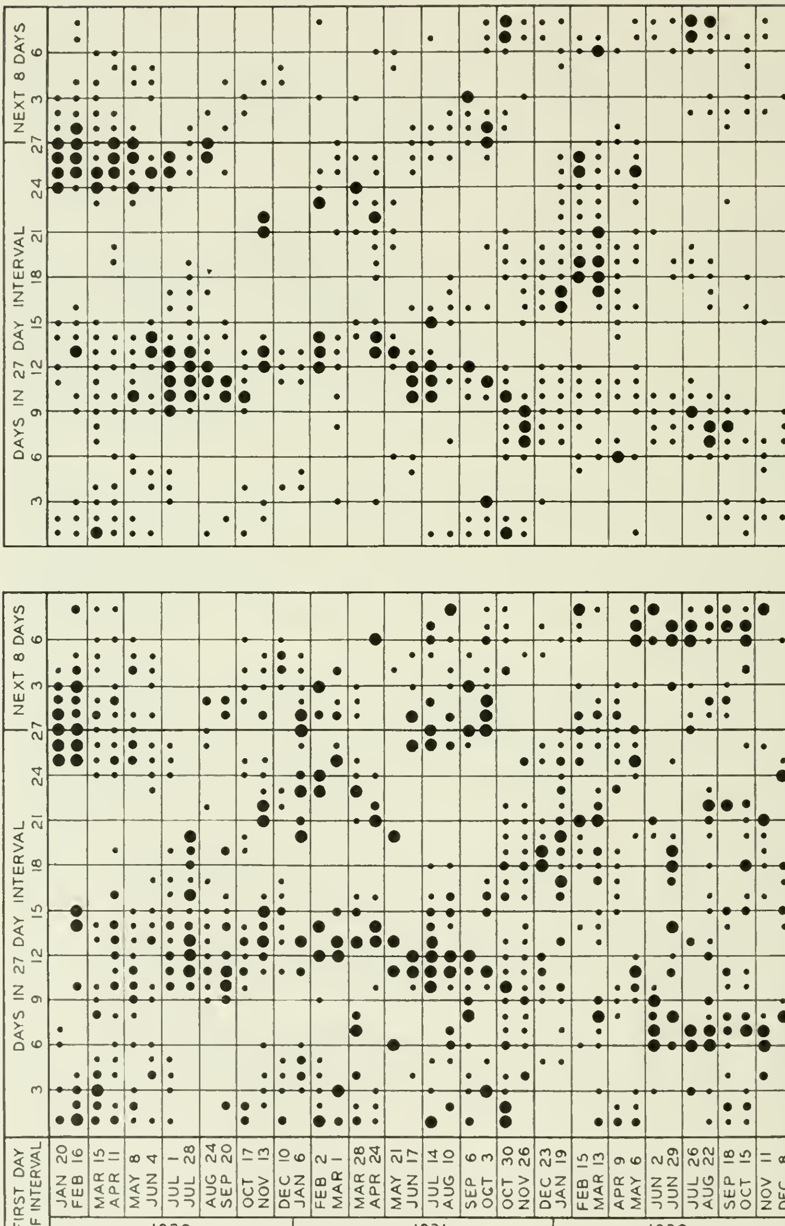


Fig. 1—Relative day-to-day record of short-wave radio transmission over the North Atlantic Ocean and terrestrial magnetism activity for 1930, 1931 and 1932, demonstrating the 27-day recurrence tendency. The dot size corresponds to the severity of disturbance.

been correct 71 per cent of the time. These figures have been determined solely on the basis of "disturbed" or "undisturbed" and modification of the forecasts by the words "probably" and "possibly" have not been taken into account.

This method of making predictions, even in its present state, is of definite use commercially. Special forecasts of the same nature have proved useful in planning certain experimental studies.

Well defined sequences of activity of an approximately 27-day period are also apparent in data of solar phenomena, particularly those relating to sunspots, bright and dark hydrogen flocculi,[1] and prominences. An attempt to link such solar sequences with the terrestrial ones noted above in a cause and effect relationship was not, however, very successful. For several well marked radio (and magnetic) sequences it was found that no single type of solar activity could be identified in such a manner as to exhibit a clear cut relationship. For some of the sequences there was some form of solar activity near the center of the sun at the time of each radio disturbance but such activity varied between recurrences in heliographic latitude and longitude and in kind.

A similar indefinite result was found by starting with the solar sequences and attempting to match the terrestrial data with them. For instance an area on the sun approximately in heliographic latitude + 10° and longitude of 315° to 329° exhibited the presence of either hydrogen flocculi, prominences, or sunspots or combinations of these on each transit across the face of the sun from October 21, 1932 to February 9, 1933, a total of five transits. Sunspots appeared in this region on the last four transits and their identity over this period of time was noted at Mt. Wilson Observatory.[2] Although the times of central meridian crossing of this area fall within a well defined sequence[3] on the radio chart (between days 6 and 9 on the left at the bottom of the chart) the absence of activity on this solar area for earlier recurrences of the radio sequences tends to vitiate the relationship between radio disturbances and those types of solar activity which were observed. Nevertheless, the reality of the 27-day period in radio is strong indication that solar activity is responsible, even though not convincingly identified in detail.

Various other criteria were used for segregating the solar data for correlation. For instance, a study of the solar distribution of flocculi

[1] Hydrogen flocculi are clouds of hydrogen gas observed with a spectrohelioscope set on the hydrogen line H_α.

[2] *Publ. Astr. Soc. Pac.*, 45, 263, 53, Feb. 1933.

[3] This sequence is considerably strengthened by increasing the number of group ndices into which the data are divided.

and spots on terrestrially disturbed and quiet days shows a maximum and minimum, respectively, about 13° west of the center (one day past); see Fig. 2. These curves are interpreted as indicating that the

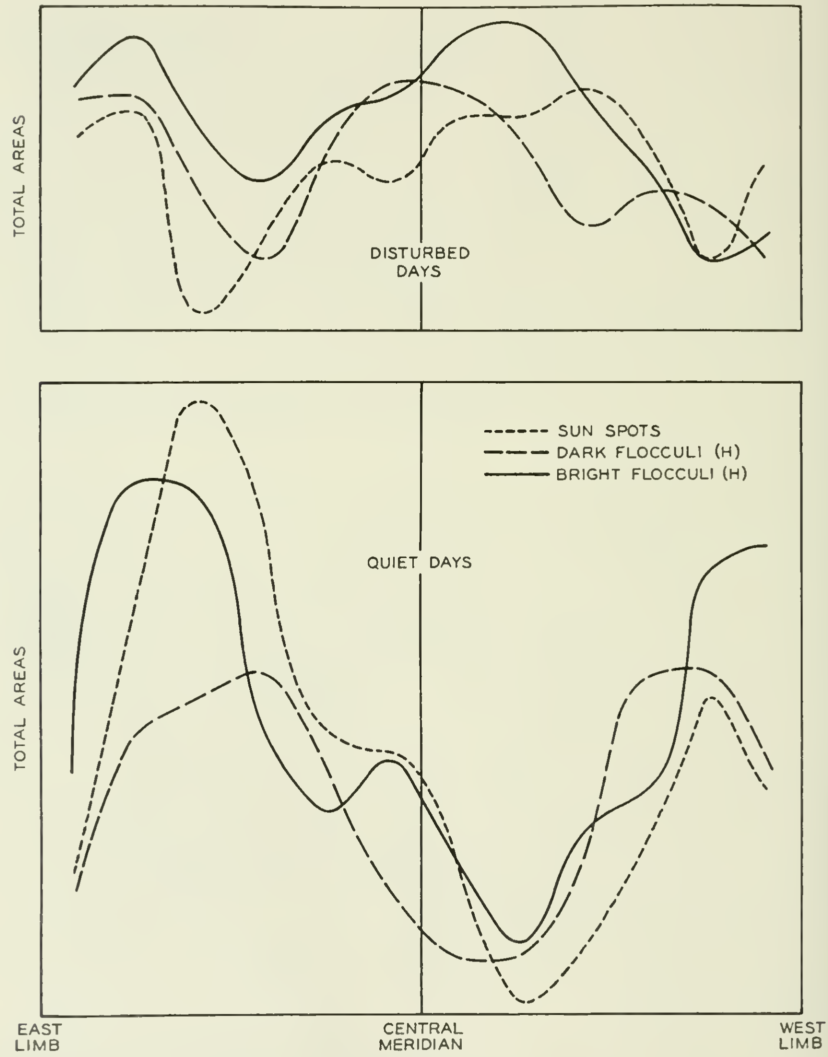


Fig. 2—Distribution across the solar disc of solar phenomena on terrestrially disturbed and quiet days.

most probable position of flocculi and spots on disturbed days is 13° west of the center of the solar disc and that on quiet days it is other

than in this region. The one-day interval from the center is interpreted as the time taken for the propagation of the disturbance from the sun to the earth.

Probably one reason for the indecisive nature of the results is to be found in the intermittent manner in which the solar data are necessarily obtained. It seems likely that considerably more success might be obtained in determining the solar-terrestrial relationships, if the solar disc could be watched continually on a world-wide program of observation, as Hale[4] has suggested, to record all solar outbursts and so to increase the completeness of the solar data until they approach those of the terrestrial.

[4] *Astrophys. Jour.*, 73, 408, 1931.

Eclipse Effects in the Ionosphere *

By J. P. SCHAFER and W. M. GOODALL

It is concluded from measurements of virtual heights and critical ionization frequencies of the various regions of the ionosphere which were made during two solar eclipses at Deal, New Jersey, that ultra-violet light is an important ionizing agency in the E, M, F_1, and F_2 regions.

As a result of pulse measurements made at Deal, New Jersey, during the partial eclipse of the sun February 3, 1935,[1] and during the total eclipse of the sun of August 31, 1932,[2] we now have data which show that the passage of the moon's shadow across the earth is accompanied by a decrease in ionization in four of the ionized regions of the ionosphere (E, M, F_1 and F_2).[3]

During the 1932 eclipse the ionic density in the E and F_1 regions was found to decrease, with the maximum effect occurring shortly after the eclipse maximum. Since the ionization in these two regions ordinarily changes uniformly with time, and since the variations observed during the eclipse were much larger than normal variations, we believe that the decrease in ionic density was actually caused by the eclipse. As regards the changes observed in the F_2 region, our 1932 results were not conclusive because the maximum effect in this region did not coincide with the eclipse but occurred somewhat later. The ionic density in this region is known to fluctuate on at least some non-eclipse days and did in fact undergo comparable variations on several occasions during the two days preceding and the two days following the eclipse. Other observers have reached the same conclusions as regards the F_2 region during the 1932 eclipse from their own data.[4]

The data from which our conclusions were drawn are shown in Fig. 1.

* Presented before joint meeting of I. R. E. and U. R. S. I., Washington, D. C., April 26, 1935. Published in same brief form in November, 1935 *I. R. E. Proceedings* as in this *Journal*.

[1] Letter to *Nature*, vol. 135, p. 393; March 9, 1935.

[2] Mention has already been made of the results of our 1932 eclipse experiments in the following publications:

Science, November 11, 1932; *Proc. Fifth Pacific Science Congress*, vol. 3, pp. 2171–2179, 1934; *Nature*, September 30, 1933; *Bell Lab. Record*, March, 1935.

The data have never been published and we are therefore including some of it in this paper as it may be of interest to other investigators in this field.

[3] M refers to the intermediate region between E and F_1.

[4] Kirby, Berkner, Gilliland, and Norton, *Proc. I. R. E.*, vol. 22, pp. 246–265, February, 1934; Henderson, *Canadian Jour. Res.*, January, 1933.

Fig. 1—Virtual height values during the total eclipse of August 31, 1932.

The double maximum in virtual height with a minimum between for 2398 and 3492.5 kilocycles [5] is interpreted by us to have been caused by a decrease in ionic density in the F_1 region, which resulted in a change in reflection from the F_1 to the F_2 layer during the central part of the eclipse.

From the curves of Fig. 1, it is possible to plot the virtual height contour map shown in Fig. 2. Since we know as a result of data taken

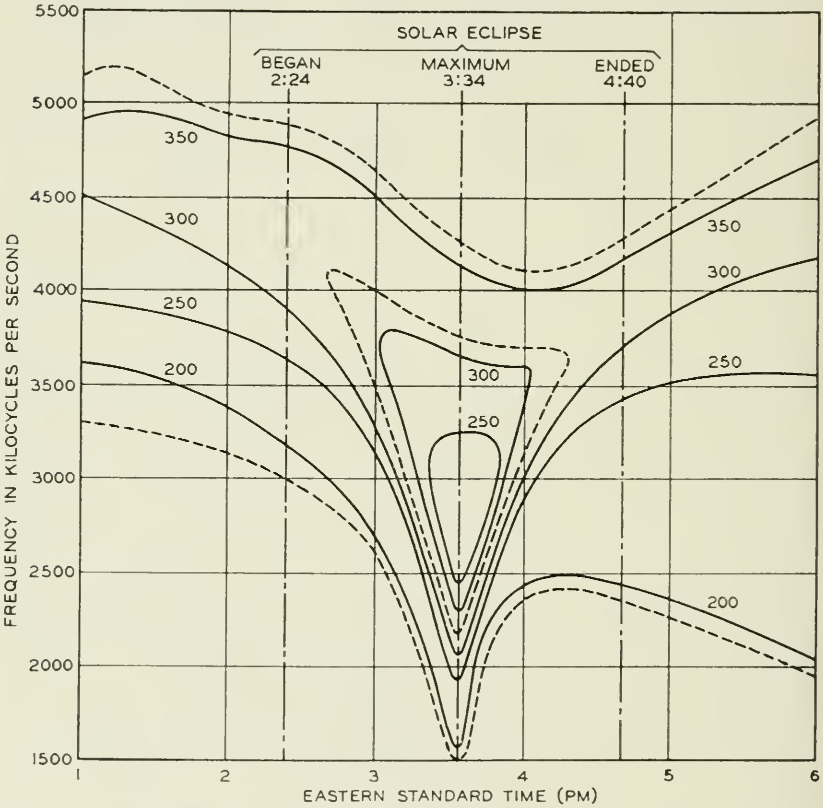


Fig. 2—Virtual height contour map drawn from data for Fig. 1, for August 31, 1932.

on several hundred days that the *X*-component curve in Fig. 1 for 4797.5 kilocycles is nearly equivalent to the *O*-component [6] curve for a frequency approximately 750 kilocycles lower (i.e., 4050 kilocycles),

[5] Mimno and Wang, *Proc. I. R. E.*, vol. 21, pp. 529–545, April, 1933, and Kenrick and Pickard, *Proc. I. R. E.*, vol. 21, pp. 546–566, April, 1933, obtained similar double maximum curves.

[6] The expressions "*O*-component" and "*X*-component" are used in place of the terms "ordinary ray" and "extraordinary ray" used by other writers.

there are in effect *O*-component curves for four different frequencies available for plotting the contour map. The dotted lines represent regions of maximum ionization. These curves show in a rather striking manner the sharp decrease in ionization of the E and F_1 regions near the time of the eclipse maximum.

The uncertainty of the 1932 results as regards the F_2 region led us to concentrate our efforts on this region during the 1935 eclipse. Improved technique now made it possible to measure the critical ionization frequencies directly instead of making virtual height measurements on fixed frequencies as had been done during the 1932 eclipse. The critical frequencies for the E and M regions were measured in addition to those for the F_2 region.

We found that this eclipse was accompanied by a decrease in the maximum ionic density in all three regions and that the minimum ionization occurred at or very shortly after the eclipse maximum. The percentage decrease was progressively greater from the lowest to the highest region, being approximately twenty per cent for the E region, twenty-two per cent for the M region, and twenty-five per cent for the F_2 region.

Some such progressive change might be expected from the fact that the eclipse had a magnitude of forty per cent at the ground and approximately forty-three per cent in the F_2 region (250 kilometers over Deal). These magnitudes are in terms of the sun's diameter, and on the basis of eclipsed area these figures become twenty-nine and thirty-one per cent, respectively.

Figure 3 gives the critical ionization frequencies for the three days on which data were obtained. The curves for the E and M regions are for the *O*-component while the curves for the F_2 region are for the *X*-component. For the *O*-component, the ionic density, N, is proportional to f_c^2, where f_c is the critical ionization frequency, while for the *X*-component curve in Fig. 3, N, is proportional to $(f_c - 750 \text{ kc.})^2$

The decreases in ionic density of the various regions may be compared with a fifty to sixty per cent decrease in the E region ionization during the eclipse of August 31, 1932, when the eclipse magnitude was ninety-five to one hundred per cent.

The 1935 measurements give a more definite synchronism than those of 1932, between the eclipse occurrence and the time of decrease in ionic density of the F_2 region.

In view of the variable nature of the F_2 region, it is a possibility that the decrease in ionic density at the time of the eclipse was a mere coincidence and was actually due to some noneclipse agency. We believe that this was not the case and that the decrease in F_2 ionization

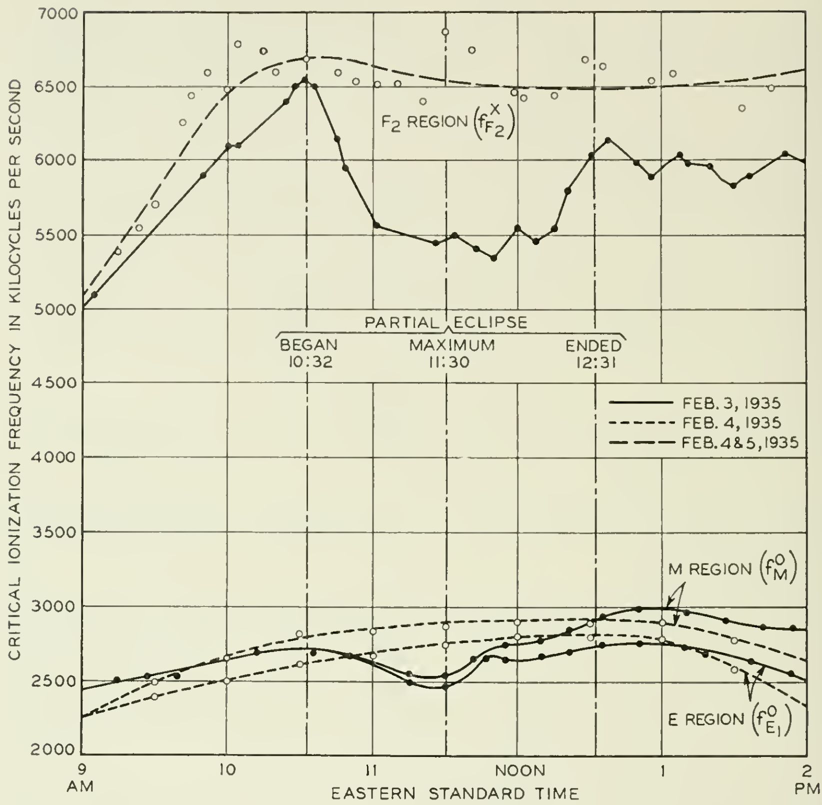


Fig. 3—Critical ionization frequencies during the partial eclipse of February 3, 1935

was a bona fide eclipse effect, as the decrease began within a few minutes after the first contact, and the density attained its lowest value shortly after the maximum of the eclipse and recovered to a more or less constant higher value a few minutes after the last contact. At no time during these measurements on the eclipse day or the days after was there any other variation of a comparable magnitude.

These results therefore indicate that ultraviolet light [7] is an important ionizing agency in the E, M, F_1 and F_2 regions of the ionosphere.

[7] While ultraviolet light is probably the ionizing agency responsible for the effects noted, any other solar emanation which travels substantially at the velocity of light, should not be precluded from consideration. *See* E. A. W. Müller, *Nature*, February 2, 1935, who suggests Roentgen type radiation.

Earth Resistivity and Geological Structure *

By R. H. CARD

IN connection with inductive coordination problems it is frequently necessary to estimate low-frequency ground-return mutual impedances between power and communication lines. The distribution of currents in the earth is a major factor in the determination of these impedances. This distribution is controlled by the resistivities of the component parts of the earth's crust and the arrangement of these parts. In impedance formulas that are customarily used the effect of the earth is taken care of by the inclusion of a single parameter—the earth resistivity. For a homogeneous earth this would be the actual resistivity of the material composing it. But the crust is nowhere homogeneous; hence, the resistivity used in such formulas is always of the nature of an average of the resistivities of the several parts of the crust—it is termed the effective earth resistivity.

The effective earth resistivities for fundamental power-system frequencies derived from mutual impedance measurements made in many parts of the world range, in general, from 2 to 10,000 meter-ohms. In a few instances values considerably higher than 10,000 meter-ohms have been observed. (The resistivity of a particular material, expressed in terms of the meter-ohm, is equal to the resistance in ohms between opposite faces of a one-meter cube of that material.)

With such a range of earth resistivities to contend with it is to be expected that estimates of ground-return mutual impedances for situations in areas where no earth resistivity data are available may be in error by large factors. In an effort to improve upon the accuracy of such estimates a study was begun several years ago of the relation between effective earth resistivity and geology. Consideration was at that time given only to areal geology, the geology of the strata of the crust lying immediately below the soil and other loose surface materials.

From this preliminary work, it appeared that the resistivities in areas of very old rocks were high and that, in a general way, decreasing resistivity corresponded to decreasing age of the rocks. There were, however, a number of outstanding discrepancies that could not be satisfactorily explained.

* Digest of a paper published in *Electrical Engineering*, November, 1935, and scheduled for presentation at the A. I. E. E. Winter Convention, New York, N. Y., January 28–31, 1936.

It then became apparent that consideration of the areal geology alone was not sufficient; instead, that the earth's structure to depths ranging from several hundred to several thousand feet must be taken into account. Data on this structure and the effective resistivities indicated by mutual impedance measurements at a large number of test sites have now been assembled. Analysis of these data shows a more or less consistent relation between the resistivity at any given point and the age and physical characteristics of the geological formations involved. This relation is such that, in general, decreasing effective resistivity corresponds roughly to decreasing age of the formations, as the earlier study seemed to indicate. However, there are certain exceptions to this rule.

The principal correlation data derived from the tests are summarized in the following tabulation. This is the result of grouping the tests in accordance with the geological periods to which the principal strata comprising the structure in each case belong and noting the ranges within which the resistivities determined by the greater part of the tests of each group lie.

Pre-Cambrian and combinations of Pre-Cambrian and Cambrian	1000–10,000 m.-ohms
Cambrian and Ordovician combinations	100–1000 m.-ohms
Ordovician to Devonian, inclusive, and combinations of these periods	50–600 m.-ohms
Carboniferous, Triassic, and combinations of Carboniferous and earlier periods	10–300 m.-ohms
Cretaceous, Tertiary, Quaternary and combinations of these periods	2–30 m.-ohms

It would be well to examine briefly the meaning of this summary and to consider its limitations. The geologists tell us that underlying the entire continent are extremely old rocks, extending downward to great depths. Little is known of the relative ages of different parts of this underlying structure. They are here grouped under the general term pre-Cambrian. In some areas pre-Cambrian rocks appear at or near the surface, the only covering being clays, soils, and other loose materials. In other areas they are overlain by rocks and sediments formed during later periods, the total thickness of which ranges up to many thousands of feet. The ages, arrangement, and characteristics of these upper strata are much better known. They are assigned by geologists to various periods in accordance with the ages during which they were formed. These periods appear in the tabulation in order from the oldest to the youngest.

In the case of tests made in areas where the pre-Cambrian rocks are overlain by younger strata it becomes necessary to consider just what portion of the structure probably influenced the test results. For

instance, at the test sites included in the second group of the summary the upper part of the structure consists of Ordovician rocks. These are underlain by Cambrian strata which, in turn, lie on the pre-Cambrian base. The question arises whether the results were influenced by the Ordovician strata alone, by both the Ordovician and Cambrian, or by the Ordovician, Cambrian, and pre-Cambrian. Calculations have been made which indicate that probably only the Ordovician and Cambrian strata were involved to any important extent. The tests in the other groups have been treated in a similar manner.

In the cases considered above the measurements were apparently influenced largely by strata of a single period or of two or more periods of about the same age. The problem is not always as simple as this. For instance, in some areas combinations of very old and very young strata occur. Areas in which the oldest rocks, the pre-Cambrian, lie directly under comparatively thin sediments of the latest periods—the Quaternary, Tertiary, and Cretaceous—are not uncommon. The effective resistivities shown by the tests in such areas range between very wide limits and the tabulation should not be taken as indicative of the values which may prevail under such conditions.

The effects of soils, glacial drift, alluvial deposits along the courses of streams, and other surface materials may also in some instances be such as to result in effective resistivities differing widely from those that would be indicated by the tabulation. The effect of local alluvial deposits, where they overlie the older rock strata, is to lower very materially the effective resistivity that would be expected were the deposits not present.

Another limitation which must be considered is concerned with the presence of rocks formed by volcanic action. Apparently such rocks usually have a high resistivity and where they occur in a comparatively young structure, the effective resistivity may be much higher than the tabulation would indicate.

The test results indicate also that the effective resistivities of structures of given periods within certain large geographical regions are markedly different from those of structures of the same periods in other large regions. Within any one of such regions, excluding areas where igneous and highly metamorphosed sedimentary rocks are involved, the effective resistivities of structures of the same period are encompassed within a comparatively narrow band.

To facilitate the use of the correlation data, the different areas within which tests have been made have been divided into groups in accordance with the geological periods of the uppermost strata within these

areas and maps have been prepared which show the boundaries of these areas and the results of the tests within them. Two such maps are shown in Figs. 1 and 2. These maps show the geological periods of the upper strata as they would appear were the overlying mantle of soil, glacial drift, local alluvial deposits and such removed. Any one

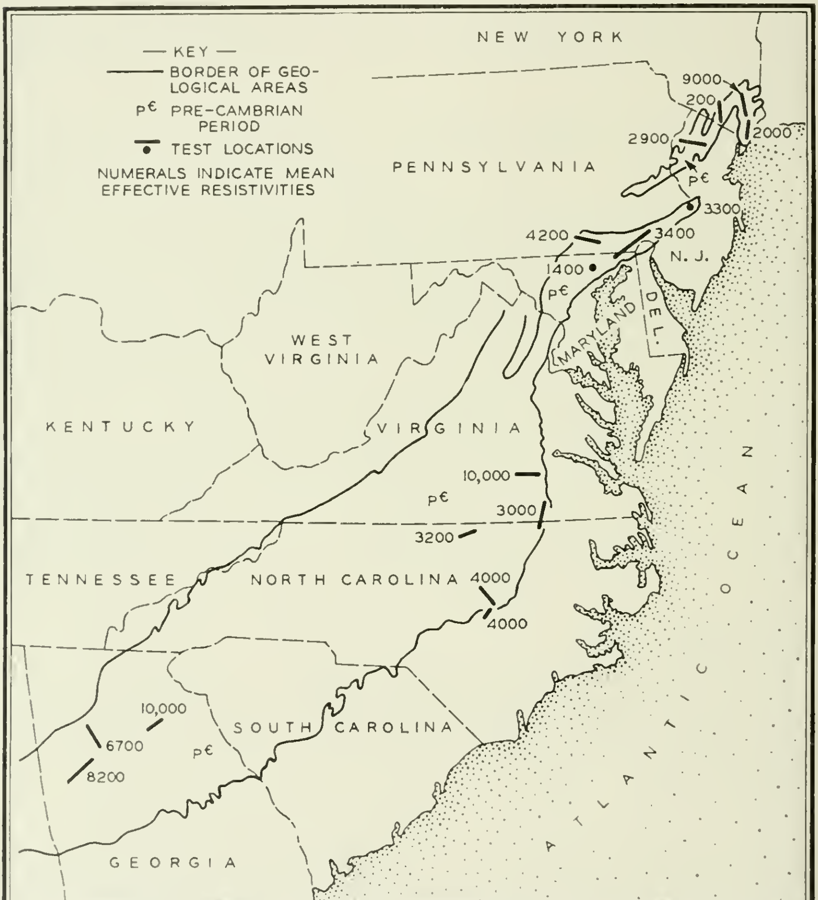


Fig. 1—Areal geology and effective resistivities—pre-Cambrian areas.

"test," as the designation is employed here, may include measurements of ground-return mutual impedances for a number of different conductors or sections of line and the test results for these different conductors or sections may indicate quite a wide range of resistivities—often 2 or 3 to 1 and occasionally 10 or more to 1, depending upon the

degree of complexity of the earth's structure at the test site. The maps show, for each test, roughly the average of the resistivities indicated by the different measurements. They also indicate for each test the relative extent of the lines involved in the test.

Within the heavy boundary lines of Fig. 1 the rocks are largely pre-Cambrian. It will be noted that most of the tests in these areas indicated very high earth resistivities—from 1400 to 10,000 meter-

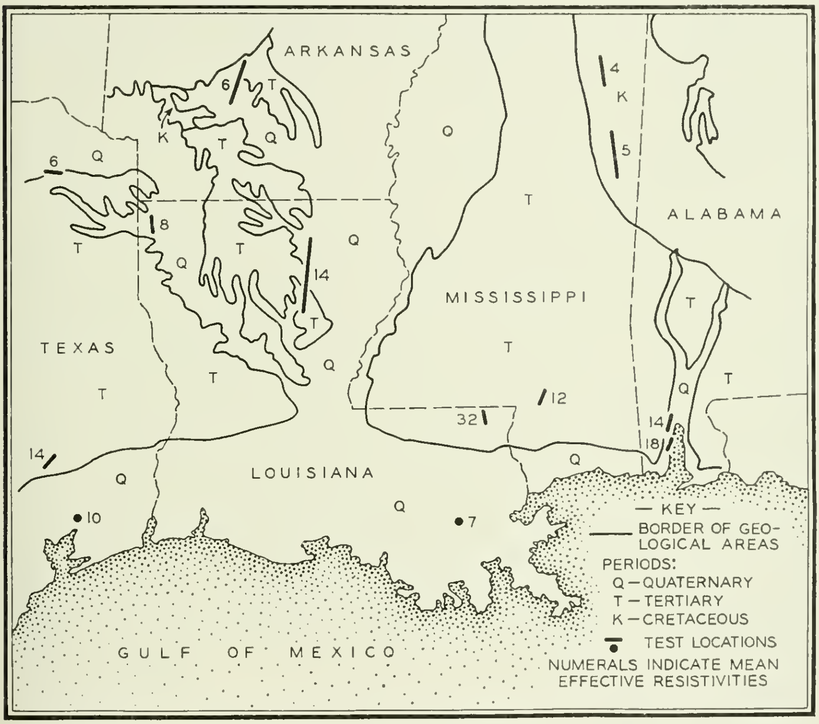


Fig. 2—Areal geology and effective resistivities—Gulf Coastal Plain areas.

ohms. The single exception, one test in New York which showed a value of 200 meter-ohms, is illustrative of the effect of alluvial deposits. The lines involved in this test were located in a narrow valley, the floor of which was partly covered with alluvium. By contrast, the structures of the areas shown in Fig. 2 are of the three latest periods, the Quaternary, Tertiary and Cretaceous, and the effective resistivities are very low—from 4 to 32 meter-ohms.

Abstracts of Technical Articles from Bell System Sources

Heat Treatment of Magnetic Materials in a Magnetic Field—II. Experiments with Two Alloys.[1] Richard M. Bozorth and Joy F. Dillinger. The magnetization of two alloys, as affected by heat treatment in a magnetic field at various temperatures, is examined in some detail in order to elucidate the nature of the accompanying changes which result in some cases in a 30-fold increase in maximum permeability. The *experiments* show that these alloys (one containing approximately 35 per cent iron and 65 per cent nickel, the other 20 per cent iron, 60 per cent cobalt and 20 per cent nickel) can be effectively heat treated in a magnetic field of 10 oersteds if the temperature is above 400° C. and below the Curie point of the alloy. The time during which the magnetic properties change has been measured at different temperatures and is found to vary according to the equation $r = Ae^{W/kT}$. The experiments are *interpreted* in terms of the domain theory of ferromagnetism. The changes which occur are due to the relief of magnetostrictive stresses which arise when the material becomes ferromagnetic upon cooling through the Curie point or when an external magnetic field is applied, and the relief comes about by plastic flow or diffusion in the separate domains. The values of A (about 10^{-12} second) and W (2.1 electron volts) are the same as those determined by Bragg and Williams for the above equation which also gives the time necessary for the establishment of a superstructure in alloys. The relation between the two processes, establishment of superstructure and the relief of magnetostrictive strains, is pointed out.

Heat Treatment of Magnetic Materials in a Magnetic Field—I. Survey of Iron-Cobalt-Nickel Alloys.[2] Joy F. Dillinger and Richard M. Bozorth. The changes that occur in the magnetic properties of iron-cobalt-nickel alloys when they are annealed in a magnetic field, have been investigated for a series of these alloys. The maximum change for the iron-nickel alloys occurs between 65 and 70 per cent nickel and is evidenced by a large increase in maximum permeability and a hysteresis loop of rectangular shape. All of the alloys with Curie points above 500° C. and with no phase transformation have their properties similarly changed. Thorough preliminary annealing

[1] *Physics*, September, 1935.
[2] *Physics*, September, 1935.

enhances the effect. With an extreme preliminary anneal of 1400° C. for 18 hours specimens of 65 permalloy have been obtained with the record value of maximum permeability of 600,000. The magnetic characteristics of materials treated in this way are relatively insensitive to stress. These magnetic characteristics are, however, highly anisotropic; the maximum permeability in one direction is as much as 150 times as large as that at right angles.

Newer Concepts of the Pitch, the Loudness and the Timbre of Musical Tones.[3] HARVEY FLETCHER. It has generally been thought that corresponding to the three psychological aspects of a sound, namely, the pitch, the loudness and the timbre, there are the three physical aspects of a sound wave, namely, the wave-length, the amplitude and the wave form. Although it is true that there is such a correspondence in a very approximate way, when the matter is examined more closely it is found that each of the psychological aspects depends upon all three of the physical properties of the sound wave.

In the paper it was shown how loudness can be defined in a quantitative way and measured by experimental methods which are described. From such measurements a relation has been found between loudness as it is ordinarily understood by the lay man and the physical intensity. In the higher intensity regions it is found that if the intensity of a sound is increased 1000-fold then the loudness will be increased 10-fold. In other words in these regions the loudness as determined by the average observer is proportional to the cube root of the intensity. For the lower intensities the loudness increases more rapidly than for the high intensities, being almost proportional to the intensity in the regions near the threshold. It is shown that the loudness depends upon the frequency, the overtone structure and the intensity of the complex sound.

In a similar way a precise definition of pitch is given which makes it possible to make quantitative measurements of this psychological aspect of a sound. Contrary to the usual notion it is found that the pitch varies not only with the fundamental frequency but also with intensity of the sound and with the overtone structure. For example, it was found that the pitch of a tone having a frequency between 100 and 200 cycles may be lowered more than a full tone by increasing the intensity without changing the frequency. Also it was shown that the pitch of a complex tone may shift as much as 1 or 2 octaves by changing the overtone structure. Numerous examples are given to show that pitch also depends upon the three physical quantities, frequency, overtone structure, and intensity. Although quantitative measurements

[3] *Jour. Franklin Institute*, October, 1935.

on timbre are still lacking there is no doubt that similar results will be found for timbre, namely, that although it depends principally upon the overtone structure, nevertheless, changes in fundamental frequency and changes in the intensity also produce large changes in the timbre.

Ceramics in the Telephone.[4] A. G. JOHNSON and L. I. SHAW. In this descriptive paper by Western Electric engineers the problems of ceramic materials as they relate to telephone usage are discussed. New products with specific properties include every type of ceramic material—electrical porcelain, vitreous enameled parts, glass and heavy clay products. Manufacturing problems of the above materials are discussed.

La Transformation Triangle—Étoile pour des Éléments de Circuits Généraux.[5] JOHN RIORDAN. This paper gives the relations between the constants of linear passive transducers connected in star and delta. The delta-star transformation previously given by Lavanchy (*Revue Générale de l'Électricité*, XXXVI, pp. 11–31 and 51–59) is shown to admit a slight generalization, perhaps of little practical importance. In the reverse (star-delta) transformation (not previously given), three of the nine independent constants of the three transducers are shown to be defined uniquely, but of the remaining six, three may be defined at pleasure, subject only to dimensional and cyclic requirements. This lack of uniqueness is shown concordant with the connection conditions.

Flutter in Sound Records.[6] T. E. SHEA, W. A. MACNAIR, and V. SUBRIZI. Frequency modulation of a sound signal is caused by non-uniformity in the record speed during the recording or reproducing process. This source of flutter is discussed and was demonstrated at the May 1935 Convention of the Society of Motion Picture Engineers.

The paper includes a discussion of the physical nature of frequency modulation, the physiological effects of frequency modulation, the methods of producing known amounts of artificial flutter, and the methods of measuring flutter.

Acoustic Impedance of Small Orifices.[7] L. J. SIVIAN. Data are presented giving the measured acoustic reactance and resistance for a number of circular orifices varying in diameter from 1 cm. down to 0.034 cm., and for a rectangular orifice 1.9 cm. $\times$ 0.075 cm. The measurements were made for various particle velocities, the cor-

[4] *Indus. and Engg. Chemistry*, November, 1935.
[5] *Revue Générale de l'Électricité*, September 21, 1935.
[6] *Jour. S. M. P. E.*, November, 1935.
[7] *Jour. Acous. Soc. Amer.*, October, 1935.

responding Reynolds numbers varying from 0.7 to 3000, roughly. The reactance is found substantially independent of the particle velocity; a formula for computing it is given. The resistance approaches a constant value as the velocity is sufficiently decreased; formulae for computing this "low velocity" resistance are given. At larger velocities the resistance increases with the velocity. This is discussed from the standpoint of a loss of kinetic energy of flow, acting besides viscosity and turbulence.

Earth-Potential Measurements Made During the International Polar Year.[8] G. C. SOUTHWORTH. Data are presented covering the normal diurnal variation of earth-potentials as measured at about a dozen different points, mostly in eastern United States. These data are arranged in graphical form for the convenience of the casual reader and also in numerical form for the use of the correlator. The data for Wyanet (Illinois), Houlton (Maine), and New York (New York) are based on nearly continuous recordings extending over a period of one or two years. This period includes the International Polar Year. At other points, less extensive data were taken. These show the general characteristics peculiar to the location in question.

The data taken at Wyanet, Houlton, and New York have been analyzed for harmonic content. At New York the fundamental and to a large extent the harmonics also, are directed along a northwest-southeast line. At Wyanet and Houlton these components tend to rotate with time. The pronounced directive effect noted near New York appears to prevail rather generally along the eastern part of the United States from Massachusetts to Florida and possibly into Cuba. The rotary effect noted in the Houlton and Wyanet data is also found in data taken in the southern part of the Mississippi Valley. Most of the data point toward the generally accepted view that there is a close relation between earth-resistivity and the direction and magnitude of earth-potentials. However, there are some inconsistencies noted which tend to make this less definite.

The Characteristics of Sound Transmission in Rooms.[9] E. C. WENTE. The characteristics of electrical circuits used for communication purposes are advantageously determined from a measurement of the transmission loss as a function of frequency. Similarly, a measurement of the acoustic pressures at various points in a room while sound of fixed intensity is emitted from a source should permit an evaluation of the acoustic characteristics of the room. Measurements of this

[8] *Terrestrial Magnetism*, September, 1935.
[9] *Jour. Acous. Soc. Amer.*, October, 1935.

type have in the past not led to any useful results because of the large variations in acoustic pressures produced by standing waves. When a high-speed level recorder is used to record pressure levels automatically as the frequency of a constant source is continuously varied, curves are obtained which not only show the variations in the general level of acoustic pressures in the various frequency regions, but which also permit an evaluation of the reverberation characteristics of different parts of the room. The paper shows curves obtained with the recorder under various room conditions.

Contributors to this Issue

Austin Bailey, A. B., University of Kansas, 1915; Ph.D., Cornell University, 1920; Instructor in Physics, Cornell University, 1915–18; Signal Corps, U. S. A., 1918–19; Assistant Professor of Physics, University of Kansas, 1921–22. American Telephone and Telegraph Company, Department of Development and Research, 1922–34; Bell Telephone Laboratories, 1934–. Dr. Bailey's work has been largely along the line of methods for making radio transmission measurements and of long-wave radio problems.

J. M. Barstow, B. S., Washburn College, 1923; M.S., University of Kansas, 1924; Instructor in Physics, Kansas State Agricultural College, 1924–27. American Telephone and Telegraph Company, Department of Development and Research, 1927–34; Bell Telephone Laboratories, 1934–. Mr. Barstow has been engaged in the development of methods and means of measuring and evaluating noise.

P. W. Blye, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1919. American Telephone and Telegraph Company, Engineering Department, 1919; Department of Development and Research, 1919–34. Bell Telephone Laboratories, 1934–. Mr. Blye has been engaged in studies of the inductive coordination of power and telephone systems from the noise standpoint.

R. M. Bozorth, A.B., Reed College, 1917; U. S. Army, 1917–19; Ph.D. in Physical Chemistry, California Institute of Technology, 1922; Research Fellow in the Institute, 1922–23. Bell Telephone Laboratories, 1923–. As Research Physicist, Dr. Bozorth is engaged in research work in magnetics.

R. M. Burns, A.B., University of Colorado, 1915; A.M., 1916; Ph.D., Princeton University, 1921; Instructor, University of Colorado, 1916–17. Second Lieutenant, Chemical Warfare Service, U. S. Army, 1918–19. Research chemist, Barrett Company, 1921–22. Western Electric Company, 1922–25. Bell Telephone Laboratories, 1925–; Assistant Chemical Director, 1931–. Dr. Burns' work has been largely in the electrochemical field and particularly on the subject of the corrosion of metals and its prevention.

Read H. Card, B.S. in Electrical Engineering, University of Tennessee, 1919. American Telephone and Telegraph Company, Long Lines

Department, 1919 and 1921–. Mr. Card's work has been concerned with transmission and inductive coordination matters.

G. W. Elmen, B.Sc., University of Nebraska, 1902; M.A., 1904; D. Engg. (hon.), 1932. Research Laboratories of the General Electric Company, 1904–06; Engineering Department of the Western Electric Company, 1906–25; Bell Telephone Laboratories, 1925–. As Research Physicist, Dr. Elmen is engaged in magnetic research.

Albert G. Ganz, M.E., Stevens Institute of Technology, 1924; M.A., Columbia University, 1931. Engineering Department, Western Electric Company, 1924–25; Bell Telephone Laboratories, 1925–. Mr. Ganz has been engaged in the development of communication coils and transformers.

W. M. Goodall, B.S. in Science, California Institute of Technology, 1928. Bell Telephone Laboratories, 1928–. Mr. Goodall has worked on ionosphere studies as well as general radio problems.

A. E. Harper, M.E., Stevens Institute of Technology, 1922. Radio Intercept Service, U. S. Army, 1918–19. Western Electric Company, Engineering Department, 1922–23; American Telephone and Telegraph Company, Department of Development and Research, 1923–34; Bell Telephone Laboratories, 1934–. Mr. Harper has been engaged in radio transmission investigations with special reference to long-wave communication.

H. E. Kent, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1923; S.M., Massachusetts Institute of Technology, 1924. National Electric Light Association, 1924–33; Edison Electric Institute, 1933–. Mr. Kent has been engaged chiefly in work relating to the inductive coordination of power and communication circuits.

Arthur G. Laird, A.B., Harvard, 1916; S.B. in Electrical Engineering, Harvard, 1921. Engineering Department of the Western Electric Company, 1921–25; Bell Telephone Laboratories, 1925–. Mr. Laird has been engaged principally in the development of communication coils and transformers.

Victor E. Legg, B.A., 1920, M.S., 1922, University of Michigan. Research Department, Detroit Edison Company, 1920–21; Bell Telephone Laboratories 1922–. Mr. Legg has been engaged in the development of magnetic materials and in their applications, par-

ticularly for the continuous loading of cables, and for compressed dust cores.

J. P. Schafer, B.S. in Electrical Engineering, Cooper Union, 1921; E.E., Cooper Union, 1925. Western Electric Company, 1915–25; Bell Telephone Laboratories, 1925–. Since 1928 Mr. Schafer has devoted a large part of his time to radio studies of the ionosphere.

S. A. Schelkunoff, B.A., M.A., in Mathematics, The State College of Washington, 1923; Ph.D. in Mathematics, Columbia University, 1928. Engineering Department, Western Electric Company, 1923–25. Bell Telephone Laboratories, 1925–26. Department of Mathematics, State College of Washington, 1926–29. Bell Telephone Laboratories, 1929–. Dr. Schelkunoff has been engaged in mathematical research, especially in the field of electromagnetic theory.

A. M. Skellett, A.B., 1924, M.S., 1927, Washington University; Ph.D., Princeton University, 1933; Instructor, 1927–28, Assistant Professor of Physics, 1928–29, University of Florida. Bell Telephone Laboratories 1929–. Dr. Skellett, formerly engaged in investigations pertaining to the transatlantic radio telephone, is concerned with applications of electronic and ionic phenomena.

OLUME XV APRIL, 1936 NUMBER 2

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

0c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each. The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.

The Bell System Technical Journal

Vol. XV *April, 1936* *No. 2*

The Reliability of Short-Wave Radio Telephone Circuits *

By R. K. POTTER and A. C. PETERSON, Jr.

From empirical measurements of noise-to-signal ratio made during the routine operation of short-wave radio telephone circuits there is obtained a general relation between percentage lost circuit time and transmission improvement in decibels. In this relation "percentage lost circuit time" is the percentage of time that the noise-to-signal ratio is considered unsatisfactory. No attempt is made to define such a standard quantitatively.

If, from past experience with a long-range, short-wave telephone, telegraph or broadcast circuit, it is known that the circuit is unsatisfactory a certain percentage of the time, the above-mentioned relation may be used to estimate the effect of transmission improvement upon this percentage of unsatisfactory or lost time. For a given circuit the variation in percentage lost circuit time, as the standard for the tolerable service is changed by a given number of decibels, may also be estimated.

There are included estimates of the relation between the number of lost time intervals of various lengths and transmission improvement.

Introduction

WITHIN a comparatively few years short-wave radio telephone circuits have become an important part of the international communication network. These years have represented a wide variety of experience ranging between the quiet and the disturbed extremes of an eleven-year sunspot cycle. An attempt is made here to review some of this transmission experience in a quantitative way and show certain relations that may be useful in the engineering of short-wave circuits.

During a magnetically disturbed year, such as 1930, a low-power short-wave transmitter with a simple antenna arrangement would have provided very uncertain means for communication across the North Atlantic. The percentage of time that such equipment could transmit what according to lenient standards in terms of noise-to-signal ratio are useful telephone signals, would have been very low. If the power of the transmitter were increased or a directive antenna employed to reduce the noise-to-signal ratio the percentage useful time would, as based upon the same standards, be increased or conversely the percentage lost time decreased. Any improvement which will decrease

*To be presented at joint meeting of U.R.S.I. and I.R.E., Washington, D. C., May 1, 1936.

the noise-to-signal ratio at the receiver output will accomplish a certain increase in usefulness of the circuit.

It is difficult to set down noise-to-signal ratios that may be employed to distinguish between satisfactory and unsatisfactory service. Such requirements would be different in the cases of telephone, telegraph and broadcast circuits. They would also depend to a considerable extent upon the facilities that it is technically and economically reasonable to provide. During times of magnetic disturbance radio telephone circuits are continued in service when noise conditions are very appreciably worse than would be tolerated on wire telephone circuits. In this emergency situation it is, of course, necessary to maintain service on the radio links as long as communication can be carried on with a reasonable degree of satisfaction.

Without a quantitative definition of the boundary between satisfactory and unsatisfactory service in terms of noise-to-signal ratio it is possible to determine from an analysis of past operating experience what percentage of the time a certain circuit was unsatisfactory. With such information available it would be useful to know how much this percentage could be reduced by the application of transmission improvements. There is developed below a form of "reliability" curve that makes it possible to estimate approximately the effect of such transmission improvements in terms of decibels upon the percentage of unsatisfactory or lost circuit time.

As a background for the following discussion it will be helpful to review briefly the conditions experienced on a typical short-wave circuit and the way in which these are related to the present analysis. For example, the instability of short-wave transmission over the North Atlantic path is well known. There are days when these transatlantic short-wave signals are remarkably good and others during times of magnetic disturbance when they are exceptionally poor. Between these two extemes is a wide range of circuit conditions. The situation is illustrated in idealized fashion by Fig. 1. The ordinates here represent average noise-to-signal ratios as measured on successive days at the receiver output and curve A of Fig. 1 (a) shows how this average might vary over an interval of many days. A certain noise-to-signal ratio such as is indicated by the horizontal line B might be specified as the highest value tolerable for a useful circuit according to some predetermined standard. Then the width of the cross-hatched intervals C represents the lost circuit time.

Fig. 1 (b) is the same as 1 (a) except that here a transmission improvement of x db has been applied so that the noise-to-signal ratio at the receiver output is on all days reduced x db and the curve A is

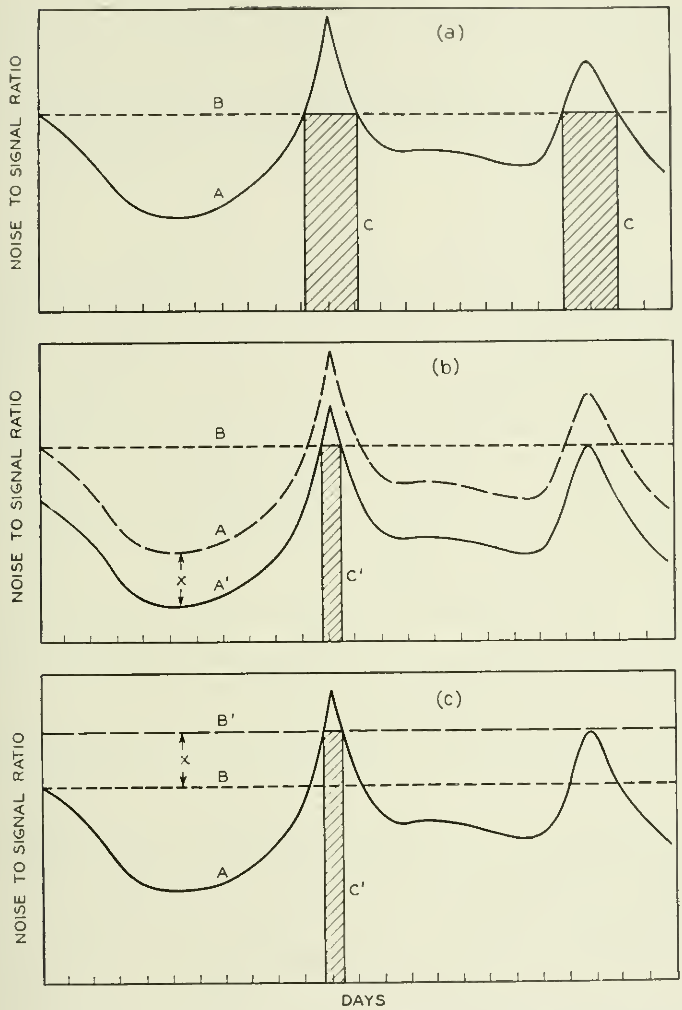


Fig. 1—Idealized illustrations of (*a*) the day-to-day variation of noise-to-signal ratio and its effect upon lost circuit time, (*b*) the effect upon lost circuit time of transmission improvements and (*c*) the effect upon lost circuit time of a change in the maximum tolerable noise-to-signal ratio.

consequently lowered to position A'. Assuming the noise-to-signal requirement B remains the same the lost time is reduced to the interval C'. If instead of applying a transmission improvement we increase the tolerable noise-to-signal requirement for a useful circuit or degrade the standard requirement by x db in the case of Fig. 1 (a), the horizontal line B is shifted upward x db as shown in Fig. 1 (c) and the lost time corresponds to that for the case of x db transmission improvement in Fig. 1 (b).

From the above illustration it is evident that if we know how the noise-to-signal ratio varies on a given circuit with a certain terminal arrangement, it is possible to determine the percentage lost circuit time for an improved or degraded system, the relative effectiveness of which can be expressed in terms of db above or below the initial arrangement.

Transatlantic Noise-to-Signal Data

As a part of the regular operating routine on the various transoceanic short-wave radio telephone circuits associated with the Bell System, measurements of noise at the receiver output are made at approximately half-hourly intervals. These measurements are made at a point in the voice-frequency wire circuits where the speech volume is normally held constant. They are therefore effectively measurements of noise-to-signal ratio although not expressed in such terms. The instrument used is known as the Western Electric 6-A Transmission Measuring set.[1] In the following discussion measurements made with this instrument are referred to as "6-A Noise."

In Fig. 2 (a) the upper curve shows the percentage distribution of 6-A noise values measured at New York during 1930 on an 18-mc. London-New York circuit. These and the curves to follow are plotted to an arithmetical probability scale. The year 1930 was severely disturbed and from the radio transmission standpoint is perhaps representative of the peak of the well-known eleven-year magnetic disturbance cycle. Since the performance of a two-way telephone circuit depends upon transmission conditions in the two directions the upper curve of Fig. 2 (a) does not accurately portray the full effect of the noise factor upon the circuit. The lower curve in this figure represents the distribution of the higher of simultaneous [2] 6-A noise values measured at New York and London. The small difference between the two distributions is evidence that the most important influence—that of magnetic disturbance—affects the transmission in both directions coincidentally.

It will be noted that these 6-A noise curves of Fig. 2 (a) and of the following figures bend downward in the region of low 6-A noise and upward where the 6-A noise becomes high. There is reason to believe that these bends are introduced by the terminal equipment and that the actual noise distribution of interest here approaches a straight line on the probability scale used, or in other words is a fortuitous

[1] L. Espenschied, "Methods for Measuring Interfering Noises," *Proc. I.R.E.*, Vol. 19, p. 1951, November, 1931.

[2] Measurements less than seven minutes apart at the two ends of the circuit were treated as "simultaneous" in this analysis.

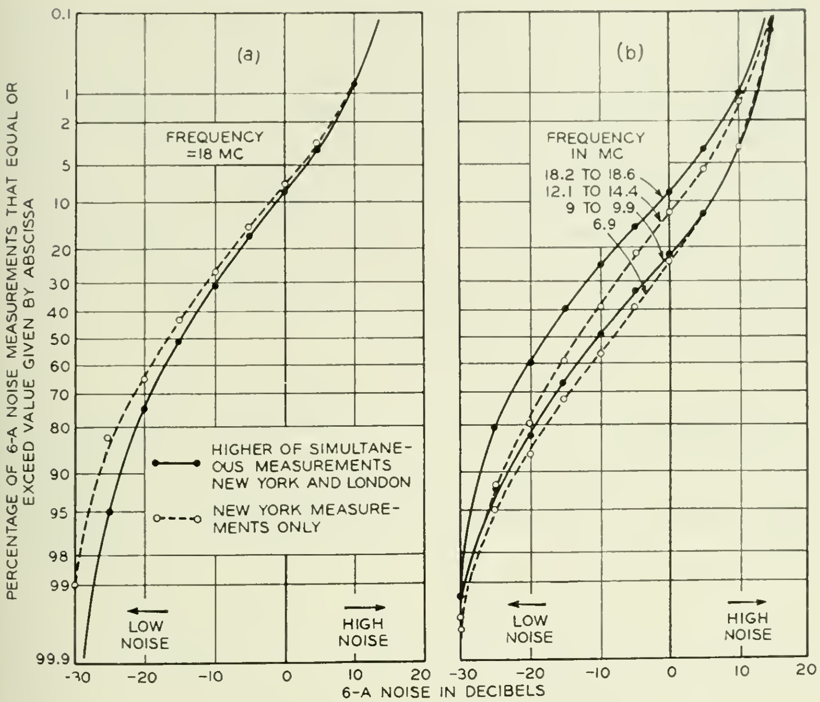


Fig. 2—Percentage distribution of 6-A noise measurements made on London to New York short-wave circuits during 1930 on (*a*) one 18-megacycle circuit and in (*b*) each of four frequency ranges.

type of distribution. The bend at the low noise end is apparently due to noise transmitted from the distant terminal and that introduced by the local receiver. These sources of noise are approximately constant and, when the atmospheric noise is very low, become the limiting factors. The bend at the high noise end of the curve is probably due to the action of the automatic volume control in the receiver. This control normally holds the speech volume approximately constant, but when noise is exceedingly high the noise in itself reduces the receiver gain and depresses both the noise and signal output. Since at such times speech volume cannot be accurately checked, the measurement is no longer an accurate indication of noise-to-signal ratio and the curve reaches a limiting value. Evidence confirming the inaccuracy of the 6-A noise readings at the high and low noise extremes will be discussed later in connection with the observed distribution of high-frequency signal intensity values.

In Fig. 2 (b) are shown distribution curves for 6-A noise values measured at New York during 1930 on several of the London to New

York circuits within the different frequency ranges indicated. All of these curves have roughly the same mid-range slope. Incidentally, if correction is made for relative transmitter power and antenna gains the horizontal separation is a direct comparison of transmission effectiveness on the different frequencies within their period of use. When such a correction is applied to the curves of Fig. 2 (b) the mid-range separation becomes less than 3 db, indicating that with equal transmitter power, antenna gains and other terminal improvements the average 6-A noise distribution is substantially the same for all times of the day when suitable frequencies are employed to cover the diurna range of transmission requirements.

Comparison of 1930, 1932 and 1934

As previously mentioned, short-wave transmission conditions were severely disturbed during the year 1930. By 1932, conditions had become much more favorable and 1934 was perhaps typical of a quiet year. In Fig. 3 (a) are included 6-A noise distribution curves for the

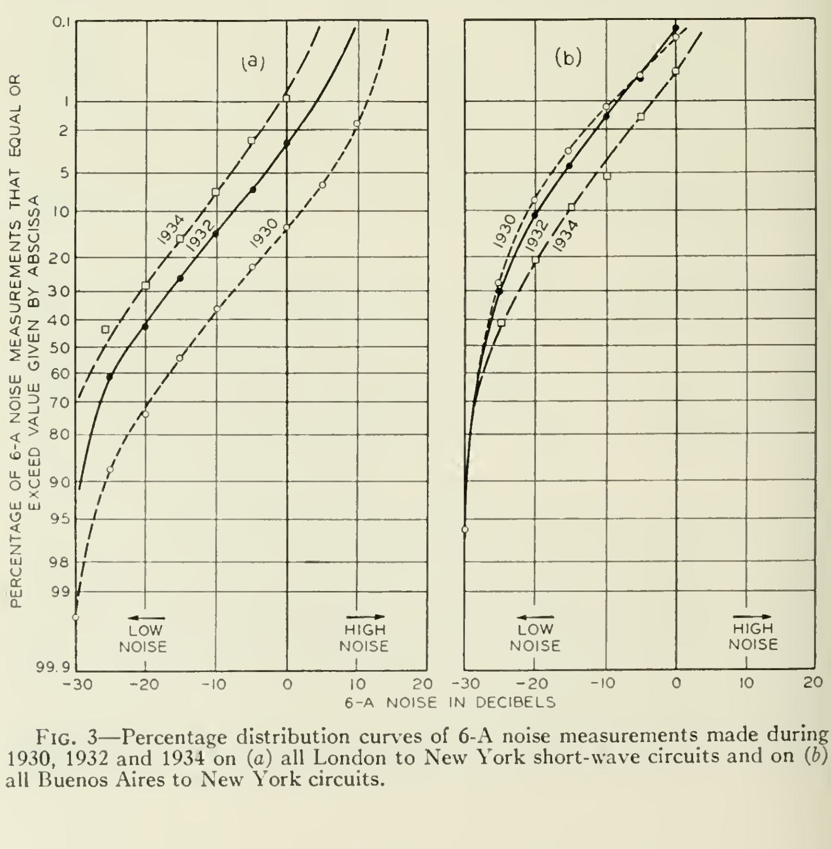


Fig. 3—Percentage distribution curves of 6-A noise measurements made during 1930, 1932 and 1934 on (*a*) all London to New York short-wave circuits and on (*b*) all Buenos Aires to New York circuits.

London-New York short-wave telephone circuits during 1930, 1932 and 1934. These three curves have very nearly the same mid-range slope, showing that the general character of the distribution was the same over the wide range of transmission conditions experienced within this interval. To those familiar with transatlantic short-wave transmission during 1930, the year 1934 would rate as comparatively undisturbed, and yet these curves indicate that for an equal percentage of measurements or, as will be apparent later, for equal lost circuit time during these two years the difference in required transmission effectiveness would be only 13 or 14 db.

Fig. 3 (b) shows that there was relatively small db separation between the 6-A noise distributions for 1930, 1932 and 1934 on the low latitude South American circuits but that the position of the curves is reversed, 1930 being better than 1934. An examination of field intensity data for these years indicates that this is due to a change in noise rather than to a change in signal transmission.

Fig. 4 (a) compares the distributions for the circuits Buenos Aires–

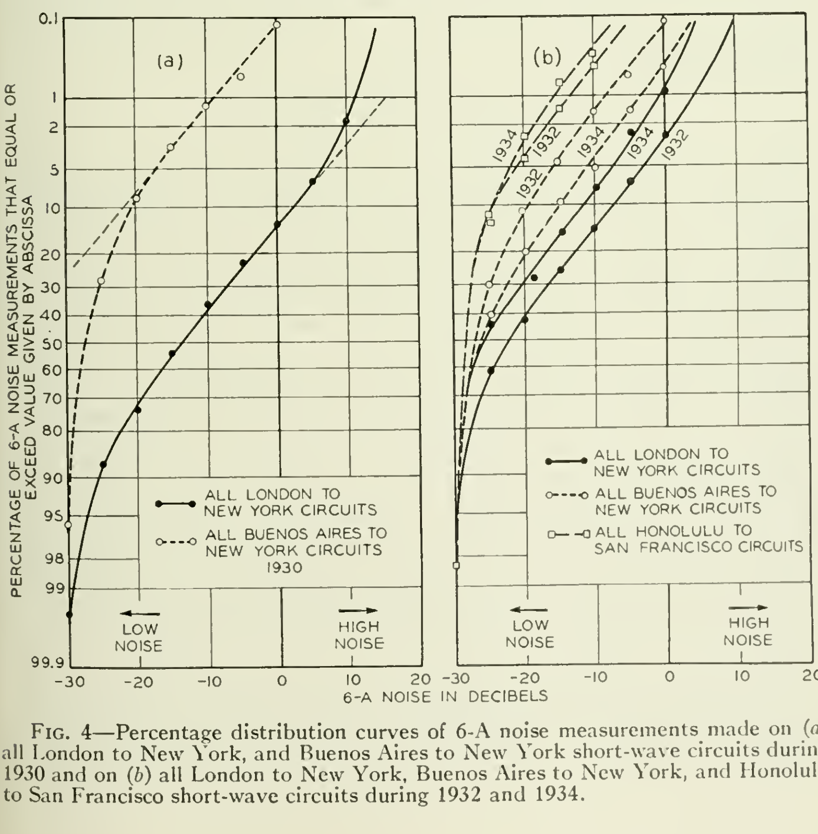


Fig. 4—Percentage distribution curves of 6-A noise measurements made on (*a*) all London to New York, and Buenos Aires to New York short-wave circuits during 1930 and on (*b*) all London to New York, Buenos Aires to New York, and Honolulu to San Francisco short-wave circuits during 1932 and 1934.

New York and London–New York for the year 1930. Again the mid-range slope as shown by the extended broken lines is about the same. The horizontal separation of roughly 25 db illustrates the much more favorable noise-to-signal conditions on the low latitude circuit, since the transmitter power and antenna gains were substantially the same in the two cases.

In Fig. 4 (b) are shown 6-A noise distribution curves for the Honolulu–San Francisco, Buenos Aires–New York and London–New York circuits representing conditions during 1932 and 1934. The mid-range slope is remarkably similar for all six curves. A comparison of these curves illustrates the high degree of reliability obtained on the circuit to Honolulu.

In Figs. 4 (a) and (b) the data for the low latitude paths cover about 9 hours of daylight operation as compared with 24-hour full time operation for the transatlantic case. Although data are not available for a 24-hour comparison some available experience indicates that such a comparison would not have altered the separation between the curves shown very appreciably. After all, the interest here is mainly in the slopes of these curves, and there is no reason to believe that the slopes would be affected.

Field Intensity Distribution

So far the discussion has shown that over the dependable portion of the 6-A noise distribution curves representative of both different circuits and different years the slopes appear to be nearly the same. If this is the case one curve may be constructed to represent approximately the effect of transmission improvement or degradation upon the performance of any long range short-wave circuit. The useful range of this curve is, however, limited by the dependable range of the 6-A noise measurements. To extend the useful range of the curve in order to estimate the effect of large changes in transmission improvement it is necessary to resort to a correction for the bend at the high noise end. Correction at the low noise end would concern the less important case of transmission degradation. Although it is possible to apply an approximate correction for the above-mentioned effect of the automatic gain control upon the bend at the high noise end it is probably more accurate to consider the distribution of field intensity data at these times of high noise-to-signal ratio.

The limiting conditions on short waves are predominantly those accompanying magnetic disturbances when the signal fields drop to very low values. The indications are that the atmospheric noise fields also decrease to a less noticeable extent during these disturb-

ances,[3] so that if this were the only effect to be considered the 6-A noise which is dependent upon the noise-to-signal ratio would increase slowly. But first circuit and tube noise in the receiver are probably the real limitation during times of disturbance. If this high-frequency first circuit noise remains constant and the field intensity decreases the 6-A noise will increase in opposite proportion. Therefore it may be assumed that the slopes of the corrected 6-A noise curves in the high noise region will correspond to the slopes of the field intensity distribution curves. In a conservative estimate it is reasonable to assume that this is the case and that although the field intensity falls during times of magnetic disturbance the high-frequency noise will not decrease.

In correcting the less important low noise ends of the 6-A noise curves, use of the field intensity distribution is not so easily justified. It may be reasoned, however, that here atmospheric noise is again low compared to receiver noise but due in this case to a scarcity of electrical storms within favorable transmission distance from the point of reception. Then the corrected 6-A noise distribution at the low noise end would also correspond in shape to the field intensity distribution. For these reasons it is assumed in the absence of better data that the field intensity distribution may be used to correct for the bends that occur at both ends of the 6-A noise distribution curves. Fairly dependable field intensity data are available over a much wider decibel range than is accurately covered by the 6-A noise measurement.

In Fig. 5 is shown by the full line *e–c–d–f* a form of noise-to-signal distribution which is conservatively representative of that experienced on several short-wave radio telephone circuits as described above. The horizontal decibel scale in this figure is arbitrarily referred to the midpoint of the distribution curve. The broken line extension *d–b* represents the decibel distribution of the lowest 15 per cent of the field intensity values as experienced during the years 1930 and 1932. The broken line extension *a–c* similarly represents the distribution of the highest 30 per cent. The reason for using the transatlantic data is that there are many more measurements available in the low field region than there are for transmission over less disturbed paths. The available data indicate that if suitable frequencies are used at all times of the day the distribution of field intensities within the lowest 15 per cent and the highest 30 per cent has roughly the same average slope during different years and on different circuits.

[3] R. K. Potter, "High Frequency Atmospheric Noise," *Proc. I.R.E.*, Vol. 19, pp. 1731–1765, October, 1931.

Circuit Reliability Curve

The corrected form of the noise-to-signal distribution curve represented by the line *a–c–d–b* in Fig. 5 may by a simple translation be put in terms of percentage lost circuit time versus transmission improvement in decibels where noise rather than quality degradation or

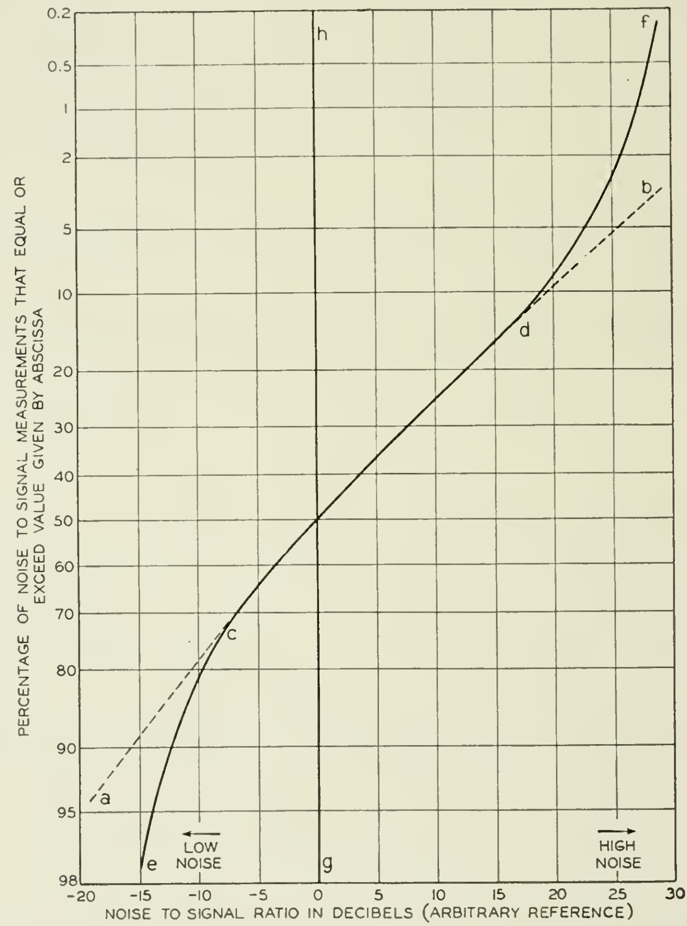


Fig. 5—Correction of assumed 6-A noise distribution curve where it is influenced by terminal equipment effects.

equipment failures interrupts continuity of service. To start this translation, take, for example, the point at which 50 per cent of the noise-to-signal values are greater and 50 per cent are less than a value given on the horizontal scale. This is 0 db and the value in itself has

no significance except as an arbitrary reference point on this scale. A vertical line is erected through this point as shown by *g–h*. If, on a certain short-wave circuit, conditions are unsatisfactory 50 per cent of the time, the effect of 10 db improvement upon this percentage may be determined by shifting the curve *a–c–d–b* of Fig. 5 ten db to the left and reading the percentage value on the vertical scale opposite the intersection of the vertical reference line. It will be remembered from previous discussion that such a shift of the 6-A noise curve toward lower values accompanies a corresponding db transmission improvement. If 50 per cent of the time conditions were unsatisfactory in the former case, they would be unsatisfactory only some 25 per cent of the time for the same tolerable noise condition and 10 db improvement. That is, the lost circuit time has been reduced from 50 to 25 per cent by 10 db transmission improvement.

By shifting the curve *a–c–d–b* of Fig. 5 various amounts to the right and left and tabulating the percentages obtained as described above, a generalized "reliability" curve may be plotted which shows the transmission improvement required to reduce the lost circuit time by any desired amount. Similarly, if we know the percentage lost time on two circuits their transmission performance may be compared on a decibel basis by determining the horizontal db separation between these two lost time values on the "reliability" curve.

A "reliability" curve of the kind described above is shown in Fig. 6. Although it is obviously unsafe to conclude on the basis of the data presented that this curve is accurately representative of all long-range short-wave circuits and circuit conditions, it serves to indicate the order of service improvement that will be afforded within the practical range of transmission improvement. For example, to reduce the lost or unsatisfactory circuit time from 50 per cent to 25 per cent appears to require about 10 db transmission improvement on any long-range short-wave circuit. Starting with a 50 per cent lost time condition and applying improvements in 10 db steps the successive percentages of lost circuit time would be roughly 25, 10, 2.5, 0.7 and 0.1.

Changes in the standards of tolerable service may be treated as equivalent to a change in the effectiveness of transmission as described earlier. Thus in terms of a high grade service the lost circuit time might for example be 50 per cent. For a grade of service 10 db lower than this the lost circuit time would be reduced to 25 per cent. The effect is equivalent to improving the transmission 10 db for the same standard of service.

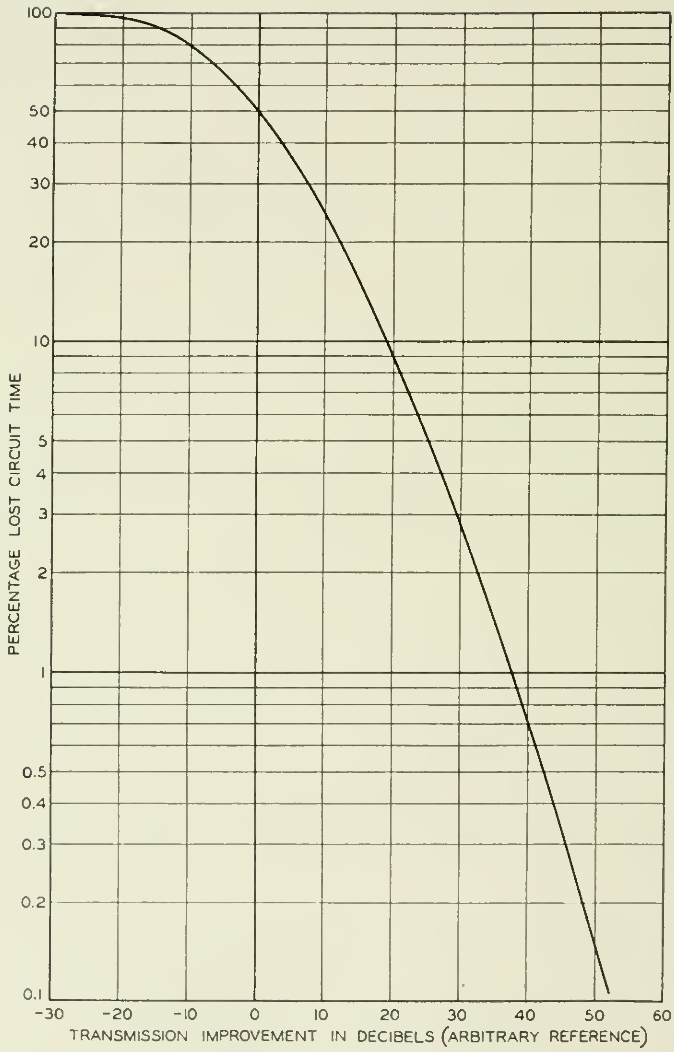


FIG. 6—Approximate relation between transmission improvement and percentage lost circuit time.

Duration of Interruptions

In the traffic operation of radio telephone circuits we are concerned in a practical way with the effect of transmission improvement upon the duration of intervals when circuits are unsatisfactory as well as upon the percentage lost or unsatisfactory circuit time. Obviously, the reduction of lost time must be accompanied by a reduction of the length of unsatisfactory intervals, or what may be termed interruptions to service. It is of interest to know how the distribution of interruptions of various lengths may be expected to vary with improvement of a circuit.

In Fig. 7 the upper curve shows how the number of circuit interruptions for the year as indicated by the ordinate value varied with hours duration on the transatlantic short-wave radio telephone circuits during 1930.[4] This upper curve was obtained from traffic data. In determining the points shown it was necessary to exclude all interruptions of uncertain length that occurred at the beginning or end of the periods when circuits were in use so that only the slope of this curve is significant.

From the field intensity measurements obtained regularly on the transatlantic circuits it is possible to obtain a useful check on the traffic experience. For example, the interruptions may be defined as the intervals of time during which no signal could be heard by beating in the carrier received on a short-wave measuring set to an audible tone with a local oscillation. By this means signals 30 db or more below those required for a barely satisfactory radio telephone circuit can be heard. In Fig. 7 the lower curve shows the distribution of interruptions based upon such a standard. This curve has substantially the same logarithmic slope as the one obtained from traffic experience. That is, the slope remains the same when the conditions defining the point of interruption are shifted by perhaps 30 or 40 db.

The summation of all interruptions shown by curves such as those in Fig. 7 should agree with the observed lost circuit time. If it is assumed, as the rather meager evidence cited above appears to indicate, that the logarithmic distribution of the interruptions remains constant for different circuit conditions, it is possible to show how the probable number of interruptions of various lengths will vary with transmission improvement. With the slope shown the position of curves corresponding to different db improvements is established by the reliability curve of Fig. 6 and the requirement that the summation of interruptions equal the lost time. Curves obtained in this manner are shown

[4] Interruptions during other years have been too infrequent to provide dependable data.

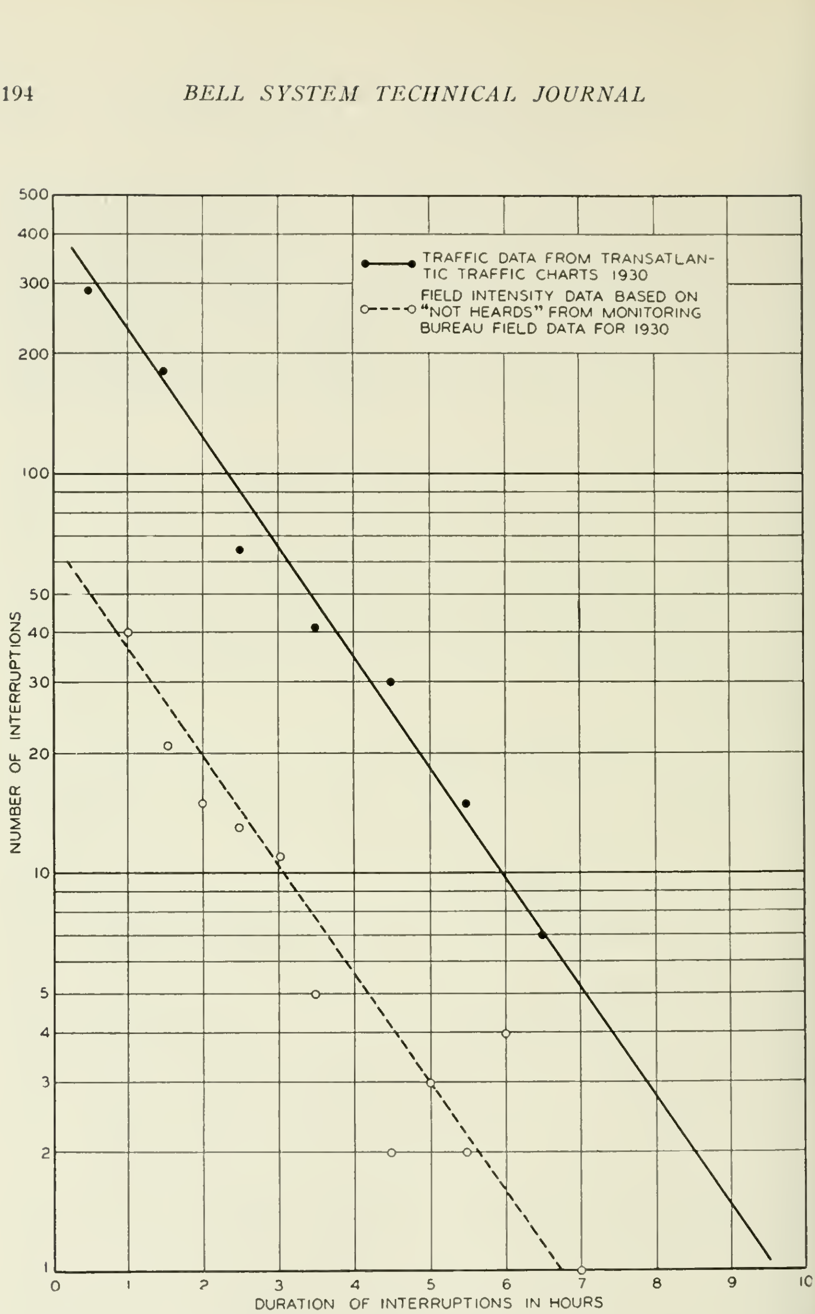


Fig. 7—Observed relation between number and duration of transatlantic short-wave radio telephone circuit interruptions during part of 1930.

in Fig. 8. As for the generalized reliability curve of Fig. 6, the reference point adopted is 50 per cent lost circuit time. Consequently the curve of Fig. 8 marked "50 per cent lost circuit time" is also designated as "0 db Transmission Improvement." Knowing the per-

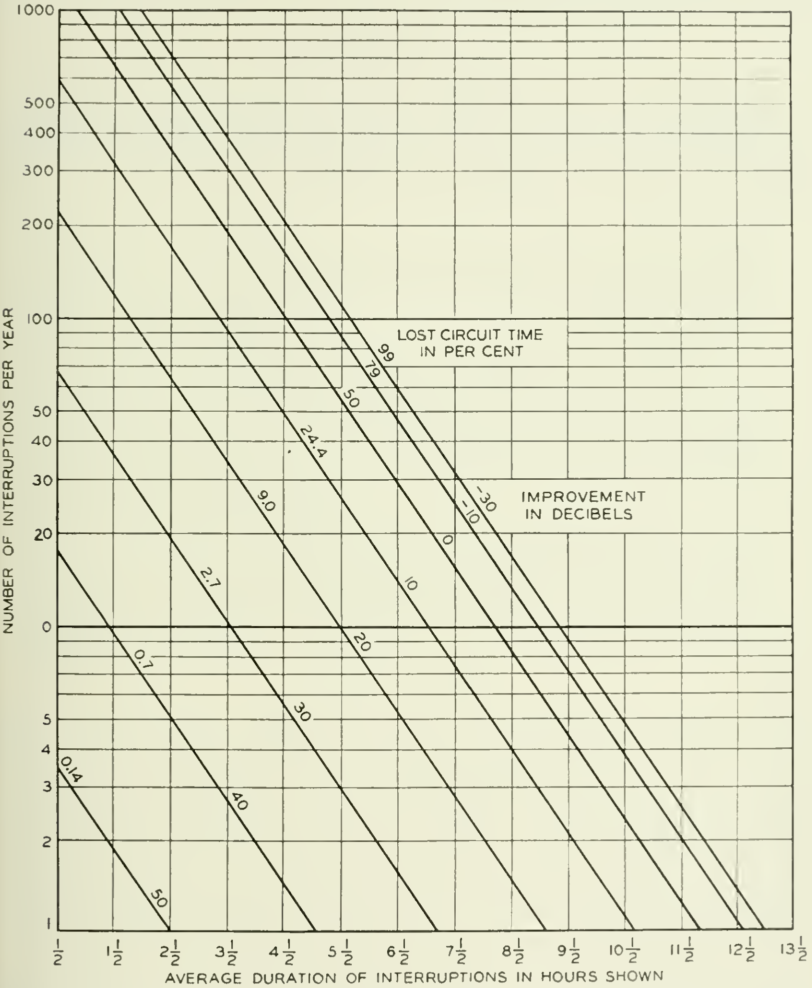


FIG. 8—An estimate of probable duration versus number of interruptions for full time operation.

centage lost time experienced on a certain circuit the position of the corresponding interruption curve in Fig. 8 may be determined by interpolation. The effect of, say, 10 db transmission improvement

upon the probable interruptions, is determined by shifting this curve 10 db to the left as measured by the indicated db spacing between curves.

It should be remembered that estimates based upon Fig. 8 are probably very approximate but the curves will at least serve to indicate the trend of improvement effects.

Acknowledgement

Acknowledgement is due the various engineers and operators of the British General Post Office and of the American Telephone and Telegraph Company who were responsible for the accumulation of the data used in this analysis.

Spontaneous Resistance Fluctuations in Carbon Microphones and Other Granular Resistances

By C. J. CHRISTENSEN and G. L. PEARSON

Voltage fluctuations which occur in resistance elements of the granular type when a direct current is flowing have been measured in the granular carbon microphone, commercial grid leaks, and sputtered or evaporated metal films. The results can be experessed by the formula

$$\overline{V_c^2} = KV^{\alpha}R^{\beta}\log(F_2/F_1),$$

where $\overline{V_c^2}$ is the mean square fluctuation voltage, V is the d.-c. voltage across the resistance R, α and β are constants having values of about 1.85 and 1.25, respectively, and F_2 and F_1 are the limits of the frequency range over which the fluctuation voltage is measured. The constant K depends, among other things on the temperature, the surrounding medium, and the dimensions and material of the resistance element; for a commonly used carbon transmitter at ordinary operating conditions its value is about 1.3×10^{-11}.

The spontaneous voltage fluctuations and the signal due to acoustic modulation are affected in almost an equivalent manner by the applied d.-c. voltage which suggests that the two effects arise from the same type of mechanism, namely a fluctuating resistance at the points of contact between granules. Experiment shows that although the acoustic signal produces a resistance modulation which is in phase at all contacts the spontaneous resistance fluctuations are completely random.

On the assumption that a region of secondary conduction, wherein the resistance fluctuation lies, surrounds each area of primary conduction as postulated in recent contact theory a value of β consistent with experiment has been deduced. On the further assumptions that thermal energy produces the mechanical fluctuations and that the equipartition law governs the distribution of energy between oscillators the observed frequency distribution follows.

Introduction

When a direct current is passed through certain types of resistance elements a small potential fluctuation between the terminals of the resistance can be observed in addition to that caused by the thermal agitation of electric charge. The resistances in which this effect is particularly noted are granular carbon microphones and commercial grid leaks which are granular in nature, such as sputtered or evaporated metal films, and any of a number of composite materials containing carbon in a finely divided state. If such a resistance element is in a current-carrying circuit associated with a telephone receiver or loud speaker, particularly when amplification is present, a steady hissing noise which sounds like that due to shot effect or thermal agitation of electric charge is heard. It is this noise which sets a practical limit to the use of the carbon microphone in sound fields of low intensity, and of commercial grid leaks in circuits carrying direct current and working at low signal levels.

The resistance of a granular conductor has been shown experimentally to lie almost entirely within very small volume elements in the regions of the contact areas.[1] It is our hypothesis that there exist minute fluctuations of resistance in the region of contact, and when such an element carries direct current a potential fluctuation between the terminals can be observed. Accordingly we propose for this phenomenon the term "contact noise."

In a study of the electrical disturbances in a carbon transmitter Kawamoto [2] found that in addition to "carbon burning," which is a sharp crackling noise sometimes present in the carbon transmitter when the voltage across individual contacts is of the order of 0.5 volt or greater,[3] there is a continuous rushing sound which is always present no matter how well the transmitter is shielded from external disturbances. Kawamoto applied the term "carbon roar" to this phenomenon. Frederick [4] in discussing the disturbances in the carbon transmitter states that the noise power is proportional to the square of the direct current passing through the transmitter. More recently Otto [5]—who has been working on this subject contemporaneously with ourselves—has reported the results of an extended investigation of this phenomenon. The present report parallels to some extent the study of Otto but in addition new aspects of the phenomenon have been investigated, more accurate data have been obtained, and the conclusions drawn from these experimental results are fundamentally different from those of Otto.

Electrical disturbance in grid leaks, which becomes evident with the passage of current, was first reported by Hull and Williams [6] who observed the phenomenon in resistances formed by an India ink line. Preliminary reports have since been published concerning such noise in thin metallic films on glass.[7] The observations of Otto [5] were also extended to fine carbon wires and copper-oxide resistances. More recently Meyer and Thiede [8] have investigated the noise in resistances consisting of thin films of carbon on a refractory base.

We have performed noise measurements on each of the types of resistance elements mentioned above and the experimental results

[1] F. S. Goucher, *Jour. Franklin Inst.* **217**, 407 (1934); *Bell Sys. Tech. Jour.* **13**, 163 (1934).
[2] T. S. Kawamoto, Unpublished Report, Engineering Division, Western Electric Company, April, 1919.
[3] This disturbance undoubtedly has its origin in the heat generated at the carbon contact by the passage of current.
[4] H. A. Frederick, *Bell Telephone Quarterly* **10**, 164, July, 1931.
[5] R. Otto, *Hochfrequenztechnik und Elektroakustik* **45**, 187 (1935).
[6] A. W. Hull and N. H. Williams, *Phys. Rev.* **25**, 173 (1925).
[7] G. W. Barnes, *Jour. Franklin Inst.* **219**, 100 (1935).
[8] Erwin Meyer and Heinz Thiede, *E.N.T.* **12**, 237 (1935).

which are presented in this paper indicate that the noise observed in each case is of the same nature and is traceable to the existence of contacts between granules or perhaps granular boundaries.

Apparatus

The experimental arrangement used in the measurements to be described here is given in schematic form in Fig. 1. The system includes the input circuit, a high gain amplifier, appropriate filters, attenuator and output measuring device.

The input circuit consists of the resistance under test, a battery for supplying the direct current, a potentiometer for measuring resistance and voltage, a standard signal oscillator for calibration purposes and appropriate resistances and condensers for coupling to the amplifier. In some cases an input transformer having a high-turns ratio was also required in order to raise the signal level above the amplifier noise level. The granular resistance element was shielded from acoustical, mechanical and electrical shocks by suspending it with rubber bands

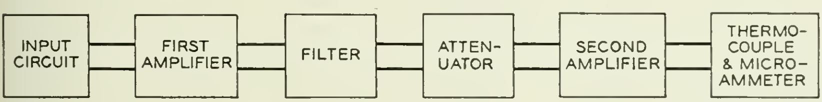


Fig. 1—Schematic amplifier circuit for measuring contact noise in granular resistance elements.

inside a tightly sealed iron box which was lined with alternate layers of hair felt and ¼-inch sheet lead. The remaining parts of the input circuit were also carefully shielded.

The high-gain amplifier consists of two separate resistance coupled units, each containing three stages. Each unit is so designed and shielded that the effect of external disturbances is eliminated. The total gain obtainable is about 165 db, with the frequency response uniform to within 2 db from 10 cycles to 15,000 cycles. In most of the measurements described here, however, a filter which transmitted only those frequencies above 100 cycles was inserted between the first amplifier unit and the attenuating network.

The gain of the amplifying system could be varied in steps of 20 db by means of interstage potentiometers. In addition, a 600-ohm attenuator having a range of 63 db in steps of 1 db was placed between the filter circuit and the second amplifier unit. The output measuring instrument was a 600-ohm vacuum thermocouple and microammeter. The deflection of the meter was closely proportional to the mean square voltage applied to the couple. Individual noise measurements were

made by adjusting the attenuation so as to bring the deflection of the microammeter as near mid-scale as possible. Fractions of a db were estimated by means of the deviation from the standard mid-scale reading. Thus what was measured in each case is the insertion loss necessary to produce a standard electrical output.

Noise as a Function of Applied d.-c. Voltage

The contact noise in several different types of granular resistance elements was measured as a function of the applied d.-c. voltage, all other variables such as resistance, frequency range, temperature, etc., being held constant. The first of these measurements to be described is that obtained by using a standard handset telephone transmitter. The circuit used for coupling to the high-gain amplifier is shown in the insert of Fig. 2, the essential parts being an input transformer having a high-turns ratio, a d.-c. voltage supply, and a standard a.-c. signal generator. The resistance of the carbon transmitter was about 50 ohms.

The results of the measurement are shown in Fig. 2 where mean square contact noise voltage is plotted as ordinate and the d.-c. voltage directly across the transmitter is plotted as abscissa, the scale being logarithmic in each case. Measurements were made as the transmitter voltage was varied from 0.00145 to 4.5 volts. This is the greatest possible voltage range in which contact noise can be observed in this instrument since the contact noise is masked at the higher voltages by carbon burning and at the lower voltages by the thermal noise of the transmitter resistance. Thus the total noise at 0.00145 volt is only slightly above thermal noise and the measured value consists of thermal plus contact noise. The two effects have been calculated separately and the latter plotted as a cross. Using this method of plotting it is seen that there is a straight line relationship between contact noise and voltage over the entire lower range. These experimental data can be accurately represented by the equation

$$\overline{V_c^2} = \text{Const. } V^\alpha, \tag{1}$$

where $\overline{V_c^2}$ is the mean square contact noise voltage, V is the d.-c. voltage across the transmitter, and α is a numerical constant having in this case the value 1.85.

By this procedure the contact noise in a number of types of carbon transmitters, filled with carbons of various origins, was measured as a function of voltage. In each case the relationship given by Eq. (1) was followed very closely over a wide range of voltages. The value of

α varied slightly from cell to cell, the extreme values being 1.75 and 1.97 with an average of about 1.85.

Figure 3 gives the results of alternate measurements of contact noise and acoustic modulation performed on a particular telephone trans-

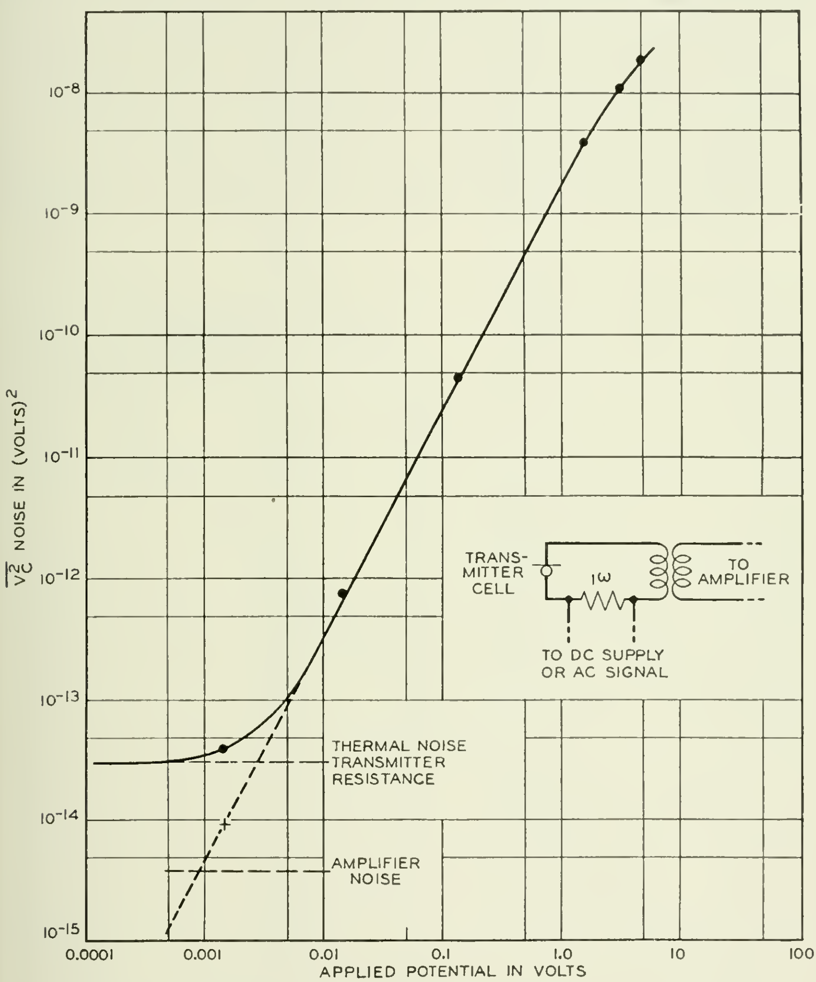


Fig. 2—The mean square contact noise voltage in a standard carbon transmitter as a function of the d.-c. voltage on the instrument. The noise level of the amplifier and the thermal noise level of the transmitter are indicated.

mitter. The acoustic field was of constant frequency and was supplied by an accurately controlled oscillator and "artificial mouth." The sound field was of such intensity that when the transmitter output was measured the background noise gave only an inappreciable part of the

whole energy. In this figure the abscissæ represent the d.-c. voltage directly across the transmitter, the scale being logarithmic, and the ordinates represent mean square contact noise or acoustic signal voltage plotted in db above an arbitrary zero level. The experimental plots for both signal and noise are straight lines but of slightly different slope.

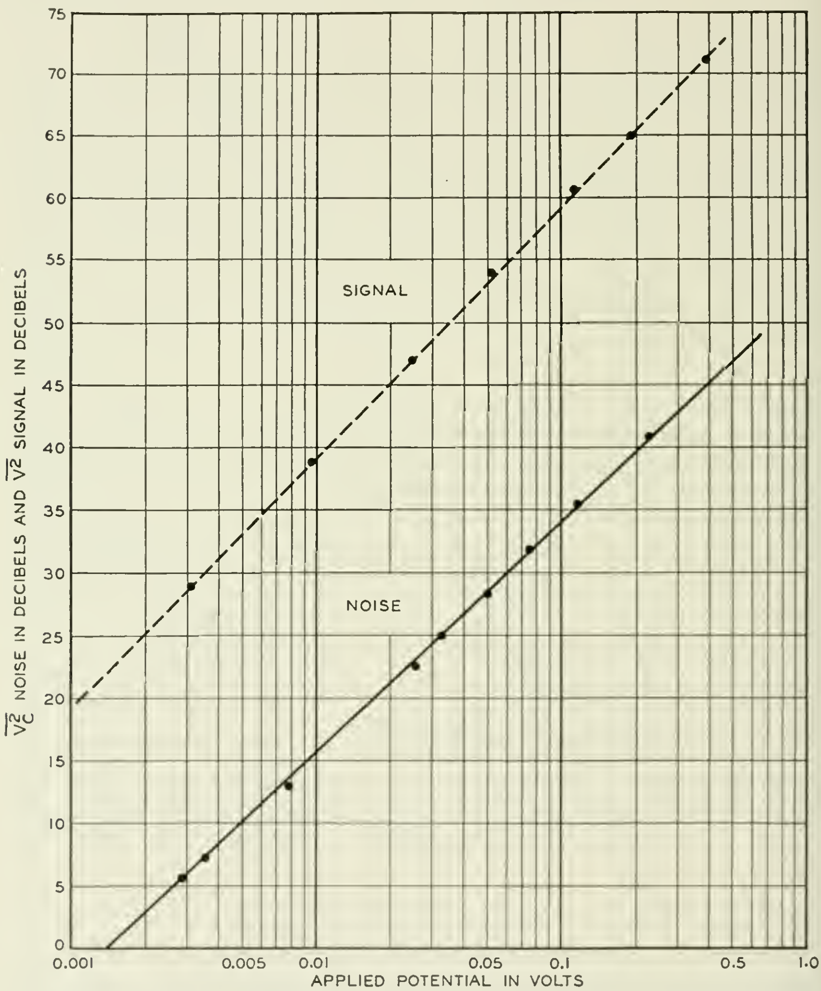


Fig. 3—The mean square contact noise and signal voltages in a standard carbon transmitter as a function of the d.-c. voltage on the instrument. The signal was obtained from a constant intensity sound field.

Analysis of the data shows that the mean square signal voltage due to acoustic modulation is accurately proportional to d.-c. transmitter voltage squared while the noise curve fixes the value of α at about 1.85. In the case of the signal the square relationship is to be expected since

the sound field produces a constant resistance modulation at the points of contact between carbon granules and the granular aggregate obeys ohms law over the entire voltage range. The fact that the noise and the signal follow so nearly the same relationship indicates that the noise also arises from resistance modulation at the points of contact between carbon granules. Since α is slightly less than 2, however, the noise mechanism is not entirely independent of the applied voltage.

Noise as a function of applied voltage was also measured in single contacts between carbon particles. For these observations a cantilever bar device was used in which the contact can be rigidly fixed and manipulated at will. This apparatus, shown in Fig. 4, consists of a

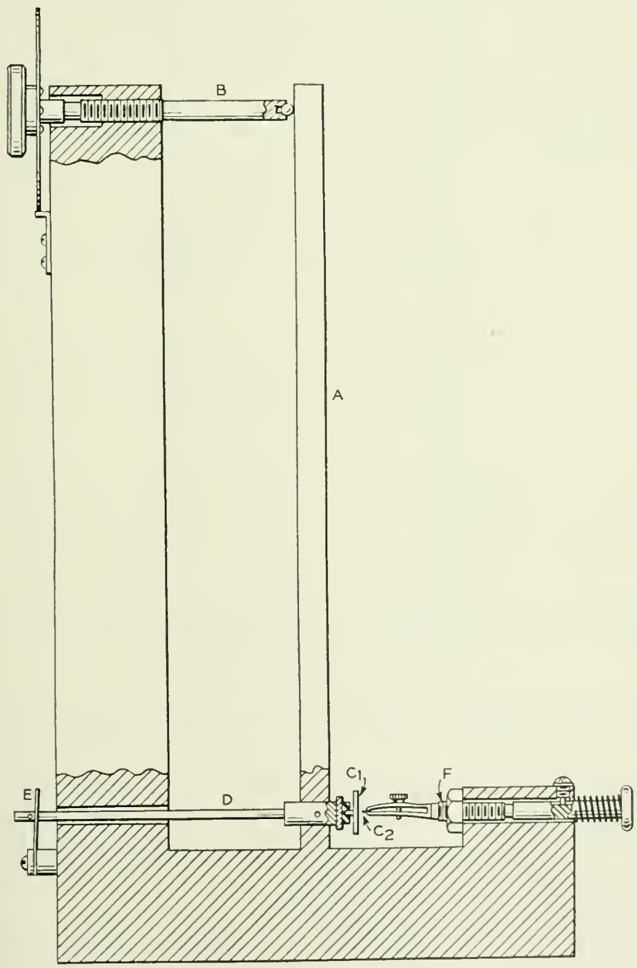


Fig. 4—Diagram of cantilever bar device for producing small contact displacements.

cantilever bar A integral with a massive L-shaped base, the entire device having been milled from a single piece of steel. A given displacement of the graduated screw B produces a greatly diminished displacement of the movable electrode C_1. The dimensions of the bar were so chosen that contact displacements of the order of 1×10^{-7} cm. could be produced. The motion of C_1 is made strictly linear by means of the pivoted rod D and any slack motion is eliminated by the

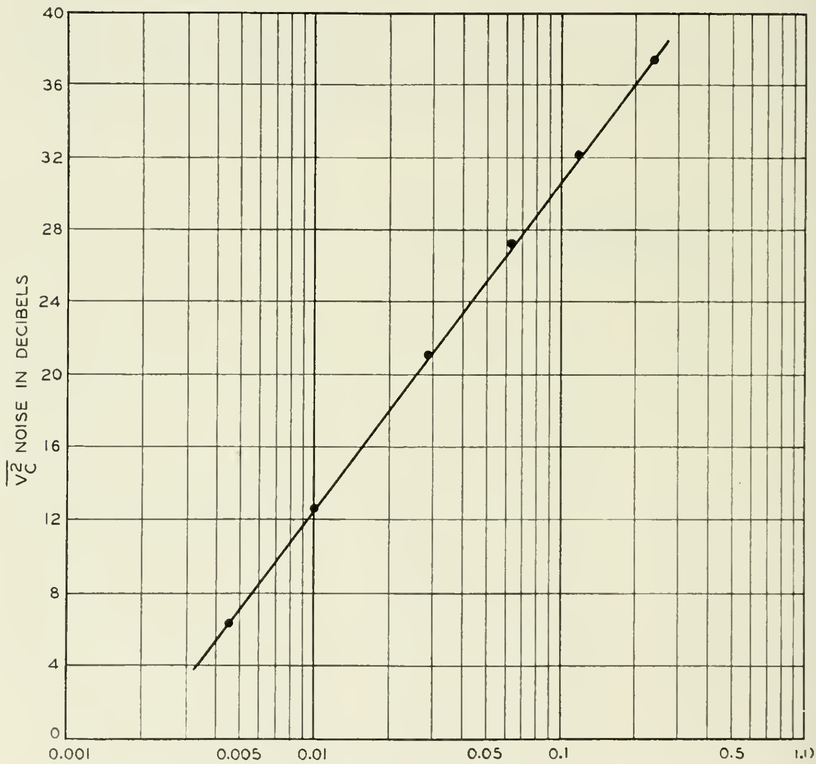


Fig. 5—The mean square contact noise voltage in a single carbon contact as a function of the applied voltage. The contact resistance was held constant at 76 ohms.

spring E. The contact is initially adjusted by the graduated screw F on which the stationary electrode C_2 is mounted. C_1 consisted of a flat polished disk of carbon like that used in the desk set transmitter, while C_2 was a composite carbon spheroid clamped securely between two gold plated jaws. Both the carbon plate and the spheroid were coated with a pyrolitic deposit of hard carbon.

The noise in a large number of single carbon contacts, connected in place of the transmitter in the input circuit shown in Fig. 2, was

measured as a function of the voltage on the contact, the resistance being held fixed. In every case the general law given by Eq. (1) was found valid, α varying between the limits 1.75 and 1.95 for different contacts. The results of a typical measurement on a contact having a resistance of 76 ohms are shown in Fig. 5. The experimental points fall on a straight line having a slope corresponding to a value of α equal to 1.83.

Figure 6 gives the results of contact noise measurements performed on a commercial grid leak which was made by coating a thin layer of

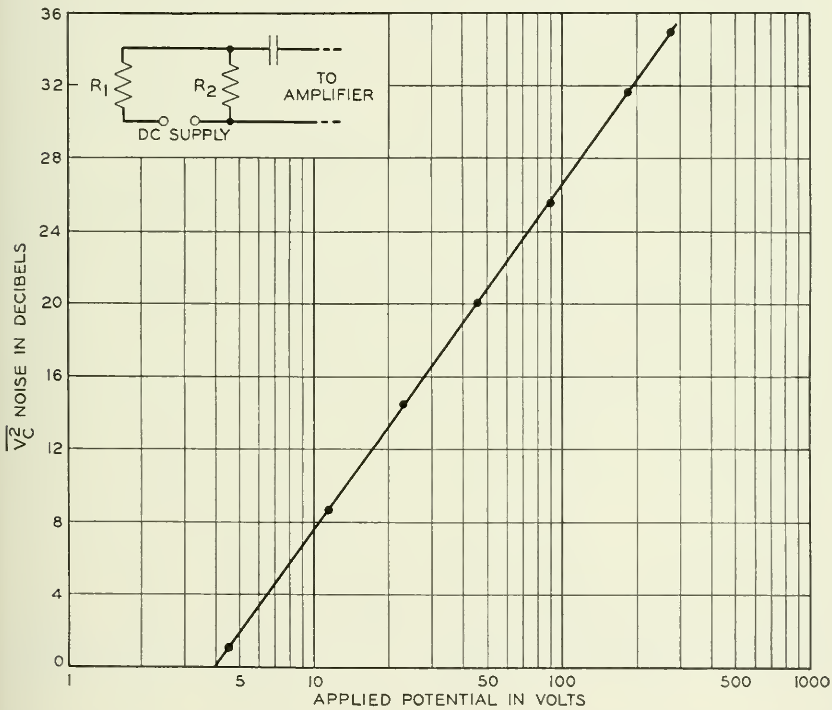


Fig. 6—The mean square contact noise voltage in a 50,000-ohm carbon grid leak resistor as a function of the applied voltage.

finely divided carbon and binder on glass. The input circuit shown in the insert, consists of R_1, the sample under test; R_2, a metal wire resistor which produces no contact noise; and a suitable source of d.-c. voltage. The resistance of both R_1 and R_2 was 50,000 ohms. The experimental points lie on a straight line the slope of which fixes the value of α at 1.90. Individual points could be reproduced within an

accuracy of 0.1 db. Similar measurements of noise in thin metallic films deposited by either the cathode sputtering or evaporation process gave results in agreement with Eq. (1), the value of α lying between the limits mentioned above.

Thus it is seen that all types of granular resistance elements which we have tested, namely, carbon transmitters, single carbon contacts and those consisting of thin films of carbon or metal follow the same relationship for noise as a function of applied d.-c. voltage.

Noise as a Function of Contact Resistance

The observation of contact noise as affected by contact resistance is necessarily limited to loose contacts, for in fixed resistance elements such as grid leaks and conducting films one has no means of independently varying their resistances. One alters the resistance of a single contact by the relative displacement of the two contacting elements. The resistance of a multi-contact device, such as a carbon microphone, may be altered either by a relative displacement of the contacting particles, or by a change in the number of contacts between the electrodes. The noise is affected differently by these two methods of resistance change; hence one must study them separately. In this section we shall be concerned only with noise as affected by resistance changes due to contact displacement both in single contacts and in aggregates; that due to a change in the number of contacts between the electrodes will be considered in our discussion of the noise from a contact assemblage.

Figure 7 is a diagram of the input circuit used to study the relation-

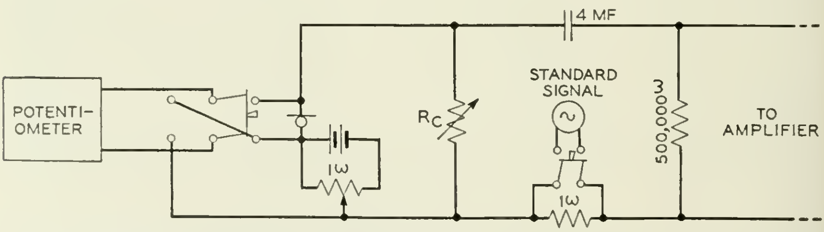


Fig. 7—Diagram of the circuit used in measuring contact noise as a function of resistance in granular resistance elements.

ship between noise and resistance. By means of the potentiometer we could measure both the voltage supplied to the contact circuit and that across the contact or transmitter. The resistance R_c is so adjusted for each noise observation that one half the voltage supplied to the circuit is across the contacts. The contact resistance for this condition is given by R_c. Also, in every case, one half the generated noise voltage is impressed on the input tube of the amplifier.

The single contacts studied were mounted in the cantilever bar device as described in the preceding section. We found it important to wait after the mounting of a contact long enough for the whole bar to come to thermal equilibrium before a measurement was attempted, otherwise very erratic results were obtained. Figure 8 is a typical curve obtained when the mean square noise voltage is plotted in db against the contact resistance on a logarithmic scale. For this

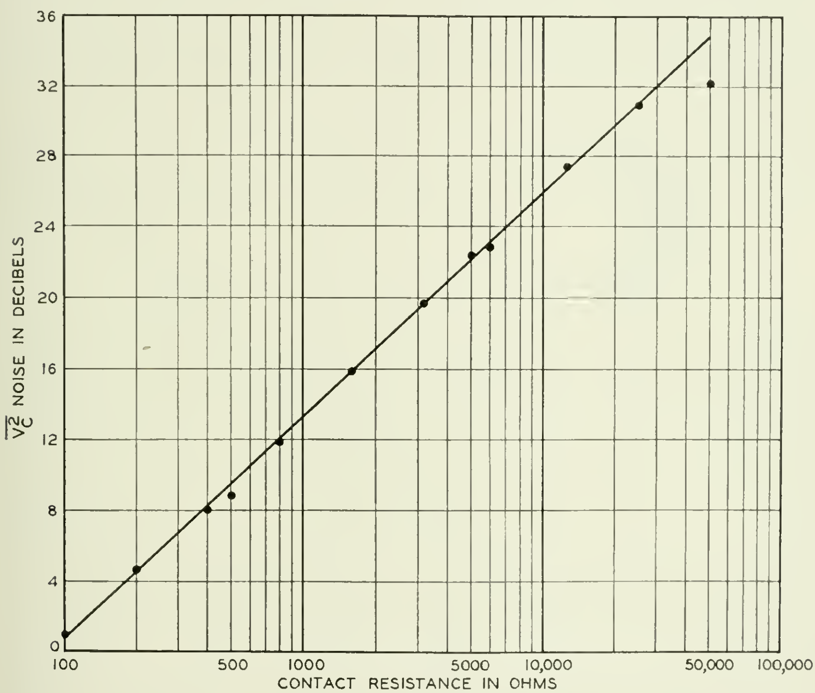


Fig. 8—The mean square contact noise voltage in a single carbon contact as a function of the contact resistance. The resistance was varied by changing the contact displacement while the applied voltage was held at 0.1 volt.

measurement the d.-c. voltage on the contact was 0.1 volt. The experimental relationship between noise and resistance is given by

$$\overline{V_c^2} = \text{Const. } R^{\beta}. \tag{2}$$

The data plotted in Fig. 8 give the value $\beta = 1.25$, which is the average value found for all the contacts studied. For individual contacts β varied between the extremes of 1.1 and 1.42. The studies on other properties of single contacts also exhibit a rather wide variability from contact to contact, hence the above result is not surprising.

A departure from the straight line relationship, such as is plotted in Fig. 8, occurs only when the contact is in a relatively high-resistance state due to a very slight compression of the contacting particles. When in this condition contacts are quite unstable and observations upon them erratic, but, in general, the noise originating in them is less than that expected if Eq. (2) held over the entire range of contact resistance values.

For the study of an aggregate of contacts the carbon cell from a barrier type transmitter was chosen. The structure of this cell (see insert Fig. 9) is such that one would expect the major portion of the

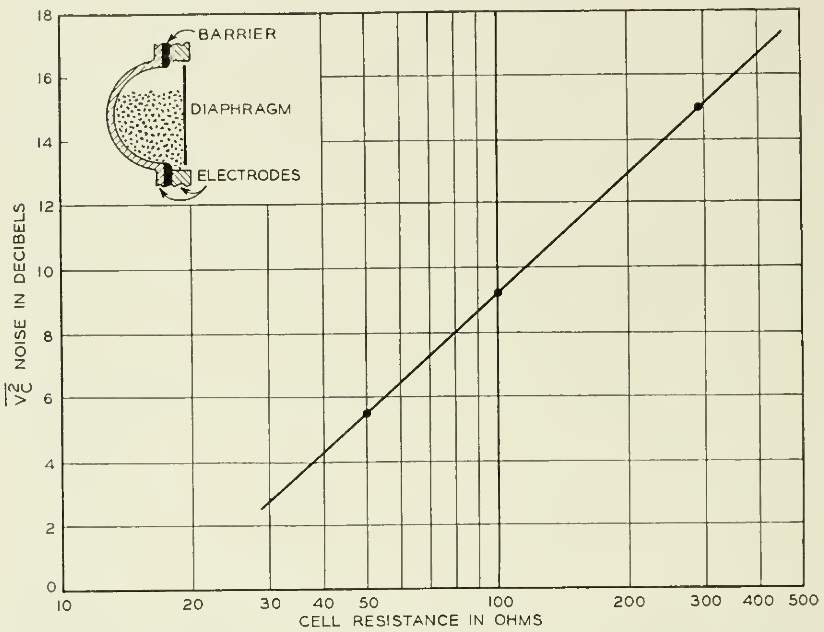


Fig. 9—The mean square contact noise voltage in a standard carbon transmitter as a function of the cell resistance. The resistance was varied by changing the amount of carbon filling while the transmitter voltage was held constant at 1.0 volt. Each experimental point represents the average of nine different readings.

current to be conducted through a small part of the total mass of granules near the bottom of the cell, and carbon added to the top of the cell to act mainly to increase the contact compressions of the conducting contacts. The electrodes of this cell are heavily gold plated, which assures that the resistance observed is almost entirely that of the contacts within the aggregate, a condition which is not fulfilled when carbon electrodes are used.

Figure 9 summarizes the results obtained in a typical set of measurements on the noise generated in the transmitter cell before mentioned. The resistance was varied by changing the height of the carbon layer above the carbon which was in the conducting path. Each of the plotted points is the average of nine observations, all of which occur in a range of 1 db. The relationship plotted in Fig. 9 can also be expressed by Eq. (2) and we again find $\beta = 1.25$. We shall see later—Eq. (8)—that when the resistance of an aggregate is altered by changing the number of conducting contacts between the electrodes quite another relationship between noise and resistance is obtained. Hence our assumption regarding the nature of the resistance change in this cell is consistent with the data of Fig. 9, and we believe that in this experiment we have measured the average value of β for all the contacts in the conducting path and have found it to be in agreement with the average value deduced from our single-contact measurements.

Noise as a Function of Frequency

For measuring the frequency distribution of the noise the filter shown in Fig. 1 was replaced by a frequency analyzer [9] having a constant band width of 20 cycles, the midpoint of which could be set at any point between 50 and 10,000 cycles per second. The calibration of the apparatus was checked by measuring the frequency distribution of thermal noise which was constant over this entire range, in accordance with theory.

The results of the measurements on a standard carbon transmitter, maintained at constant resistance and applied voltage, are shown in Fig. 10 where ordinates represent the mean square noise voltage over the 20-cycle band and abscissæ represent the mid-frequency of the band. It is seen that the experimental points fall on a straight line having a negative slope of about 1.0. This relationship may be represented by the equation

$$\Delta \overline{V_c^2} = \text{Const. } \Delta F/F, \tag{3}$$

where $\Delta \overline{V_c^2}$ is the mean square noise voltage for the frequency band ΔF. Integrating Eq. (3) between fixed limits we obtain:

$$\overline{V_c^2} = \text{Const. } \log (F_2/F_1), \tag{4}$$

which gives the total noise over the frequency range F_1 to F_2.

Figure 11 gives the results of similar measurements on a high-

[9] T. G. Castner, *Bell Laboratories Record* **13**, 267 (1935).

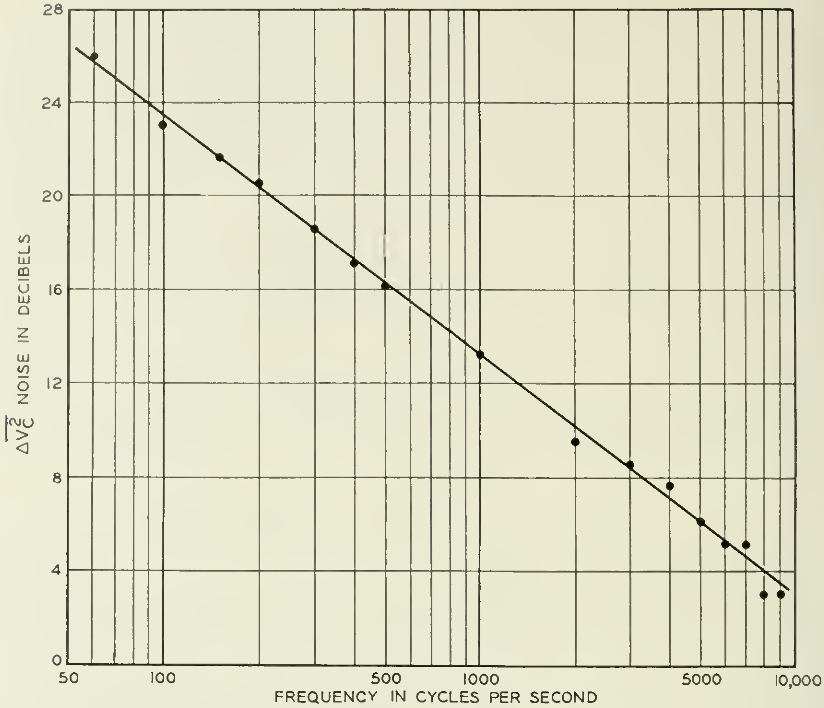


Fig. 10—The frequency distribution of contact noise in a standard carbon transmitter for normal operating conditions. The ordinates give the mean square contact noise voltage in a twenty-cycle band the midpoints of which are given by the abscissæ.

resistance carbon grid leak. It is seen that the noise has precisely the same frequency distribution in both the carbon transmitter and in the grid leak, which further supports our belief that the noise mechanism is the same in each case.

Otto [5] reported similar measurements of noise as a function of frequency in carbon transmitters, single contacts of carbon, carbon grid leaks and copper oxide resistances. Whereas we find an almost exact inverse relationship between noise and frequency for all types of elements tested he shows curves with negative slopes ranging from 1.0 to 1.4. Meyer and Thiede [8] in measurements on thin carbon films obtained negative slopes having values between 1.0 and 2.0.

Contact Noise as a Function of Temperature and Surrounding Medium

For the complete elucidation of contact noise the knowledge of its dependence upon temperature is important. However, the difficulties involved in such a measurement are so great that we have been unable,

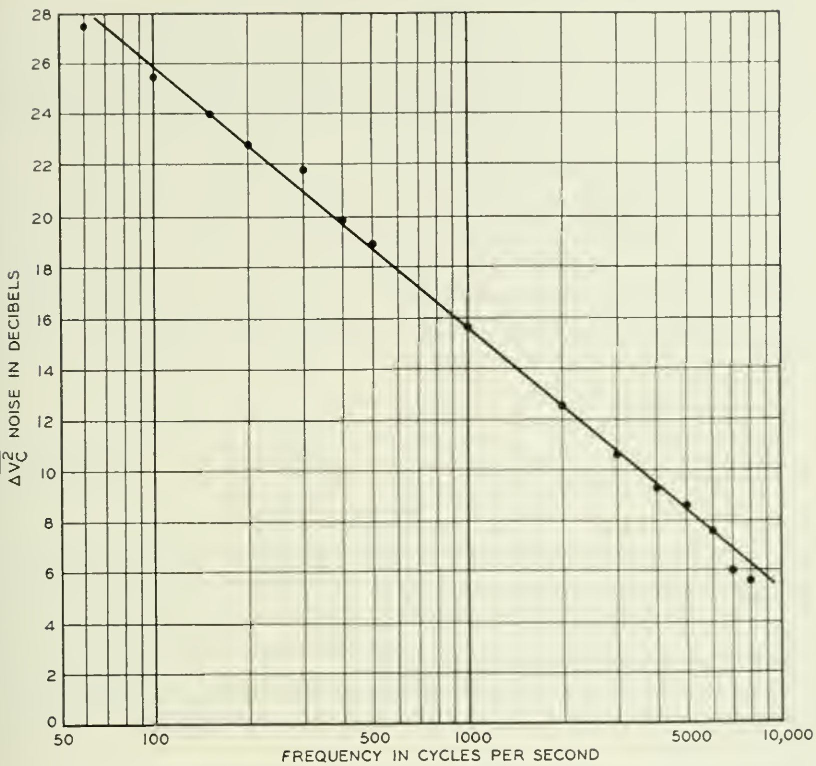


Fig. 11—The frequency distribution of the contact noise in a 50,000-ohm carbon grid leak resistor. The method of plotting the experimental points is the same as that shown in Fig. 10.

as yet, to obtain dependable results. For a satisfactory observation of the effect of temperature on the noise of contacts one must be certain that the conducting area in a contact remains constant and independent of temperature. Temperature variations can alter the conducting area in at least two ways; the contacting particles may be relatively displaced due to differential thermal expansions of the apparatus, and the change of the quantity of adsorbed gas on the contacting surfaces can alter the contacting areas without any relative displacement of the contacting particles. Both of these conditions are very difficult to control in any measurement involving temperature changes. However, our measurements of the relationship between contact noise and temperature, performed under the most carefully controlled conditions which we have been able to apply, indicate that contact noise may change either positively or negatively as a function

of temperature depending upon the conditions of the experiment, but that the total change is not more than 3 or 4 db in the temperature range from 90 to 300 degrees Kelvin. These observations are consistent with the fact that Otto [5] has reported a decrease of noise with increase of temperature while Meyer and Thiede [8] have reported the reverse effect.

The nature of the surrounding medium seems to affect the intensity of the noise but little. The noise of the contact under oil seems to be slightly less, one or two decibels, than when the contact is in a vacuum of 10^{-5} mm. of mercury. The noise in air seems to be intermediate between these two extremes. This leads us to believe that the noise mechanism is not associated with the medium surrounding the contact.

Quantitative Values of Contact Noise

Equations (1), (2) and (4) may be combined to give the expression

$$\overline{V_c^2} = K\ V^\alpha R^\beta \log\,(F_2/F_1). \tag{5}$$

This is the general empirical equation found for noise in granular resistance elements as a function of voltage, resistance, and frequency. The average values for α and β are respectively 1.85 and 1.25. The constant K is dependent on the material, shape, temperature, etc. of the resistance element. The following representative values of this constant were obtained for some of the resistance elements which we have measured:

Single carbon contact	1.2×10^{-10}
Western Electric No. 395-B telephone transmitter	1.3×10^{-11}
100,000 ohm carbon grid leak	1.1×10^{-21}

For different single carbon contacts this constant did not vary more than 20 per cent as long as a given type of carbon was used, and a change in the type of microphonic carbon produced a variation by not more than a factor of two.

The contact noise in a Western Electric No. 395-B telephone transmitter under actual working conditions (R = 45 ohms, V = 2.5 volts, F_1 = 200 c.p.s. and F_2 = 3000 c.p.s.) is given by Eq. (5) as 9.8×10^{-5} volts. The output signal of this transmitter for standard voice operation is about 0.1 volt. The spread between signal and contact noise is so great that this noise is not a disturbing factor in the standard carbon transmitter as used in telephone service. This is not true, however, in the case of high quality carbon transmitters used for studio work and public address systems. The sound fields under the conditions wherein such instruments are apt to be used are much less intense

than for the telephone transmitter under its normal condition of use, and hence, the contact noise becomes a limiting factor when the carbon microphone is used in weak sound fields.

Equation (2) is not suited for representing contact noise as a function of resistance in grid leaks since a change in resistance is brought about by variations in the dimensions and materials of the conducting film rather than by a change in contact displacement as is done in the case of loose contacts. For this reason the constant given above applies only to 100,000-ohm resistances of a given type.

It is of interest to note that the constant for the carbon grid leak resistor is smaller than that for the single contact by a factor of 10^9. This is due, in part, to the fact that the total voltage V across the grid leak resistor is divided among a network of contacts each of which produces noise independently of the others. The total noise from such an assemblage, as will be shown in the following section, is less than that arising from a single contact. This suggests that the contact noise in a solid carbon filament, if it exists at all, should be still smaller than that in the grid leak. We have made measurements on such a filament having a diameter of 0.0025 cm. and a resistance of 75,000 ohms. After taking great precautions to eliminate all the noise at the terminal connections we were unable to detect any noise in addition to that of thermal agitation for d.-c. loads as great as the filament would carry without being destroyed (a current density of 3×10^3 amperes per square cm.).

Noise From a Contact Assemblage

A transmitter cell contains an assemblage of contacts and we have shown that the noise from such an assemblage follows the empirical law set forth in Eq. (5), which also holds for single contacts. Several important deductions are possible when we study the noise from an assemblage as a function of the number and arrangement of the contacts within it.

Assemblage With Contacts in Parallel

Consider n contacts, $R_1, R_2 \cdots R_n$, placed in parallel across a direct current supply and inductance as shown in Fig. 12A. The inductance is large enough so that it offers an effectively infinite impedance to the fluctuation voltage we expect to study. Due to the fluctuation of resistance in the contacts and the passage of direct current they will act as a.-c. generators. Figure 12B is the equivalent a.-c. circuit, where $e_1, e_2 \cdots e_n$ are the instantaneous a.-c. voltages generated because of the fluctuating resistance in the respective contacts. The instantane-

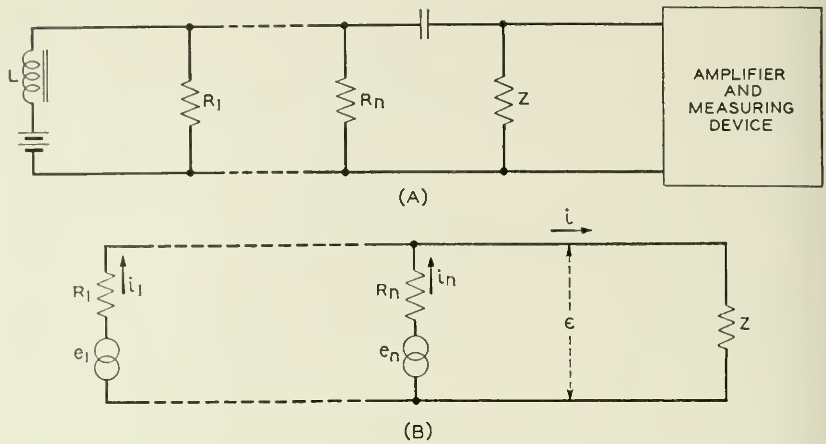


Fig. 12—(*A*) Circuit for measuring the contact noise of a parallel assemblage of resistance elements. (*B*) The equivalent a.-c. circuit of (*A*).

ous value of a.-c. voltage experienced across the impedance Z—which is the input impedance of a measuring circuit—is

$$\epsilon = e_1 - i_1 R_1 = \cdots e_n - i_n R_n = iZ,$$

where $i_1, i_2 \cdots i_n$ are the respective fluctuating currents flowing because of the generator action of the fluctuating contact resistances. Also we have

$$i = i_1 + i_2 \cdots i_n.$$

From these two expressions we get

$$\frac{e_1}{R_1} + \frac{e_2}{R_2} + \cdots \frac{e_n}{R_n} = \epsilon \left[\frac{1}{R_1} + \frac{1}{R_2} + \cdots \frac{1}{R_n} + \frac{1}{Z} \right] = \frac{\epsilon}{Y},$$

where

$$\frac{1}{Y} = \frac{1}{R_1} + \cdots \frac{1}{R_n} + \frac{1}{Z}.$$

Hence we obtain

$$\epsilon = e_1 \frac{Y}{R_1} + \cdots e_n \frac{Y}{R_n}.$$

The noise power dissipated in Z, and thus measured by the measuring device, is defined in the usual way as $\overline{\epsilon^2}/Z$, where $\overline{\epsilon^2}$ is the mean square voltage across the impedance Z and is defined as

$$\overline{\epsilon^2} = \underset{t \to \infty}{\text{limit}} \, (1/t) \int_0^t \epsilon^2 dt.$$

If the value of ϵ as given above is substituted in this equation, and the relative phases of the individual contact voltages $e_1 \cdots e_n$ are considered random, one derives the relationship

$$\overline{\epsilon^2}_{\text{parallel}} = \frac{\dfrac{\overline{e_1^2}}{R_1^2} + \dfrac{\overline{e_2^2}}{R_2^2} + \cdots \dfrac{\overline{e_n^2}}{R_n^2}}{\left[\dfrac{1}{R_1} + \dfrac{1}{R_2} + \cdots \dfrac{1}{R_1} + \dfrac{1}{Z}\right]^2}. \tag{6}$$

The experimental test of this derived relationship will be given later.

Assemblage with Contacts in Series

Consider n contacts placed in series and supplied with current from a battery through an inductance as shown in Fig. 13*A*. The equivalent

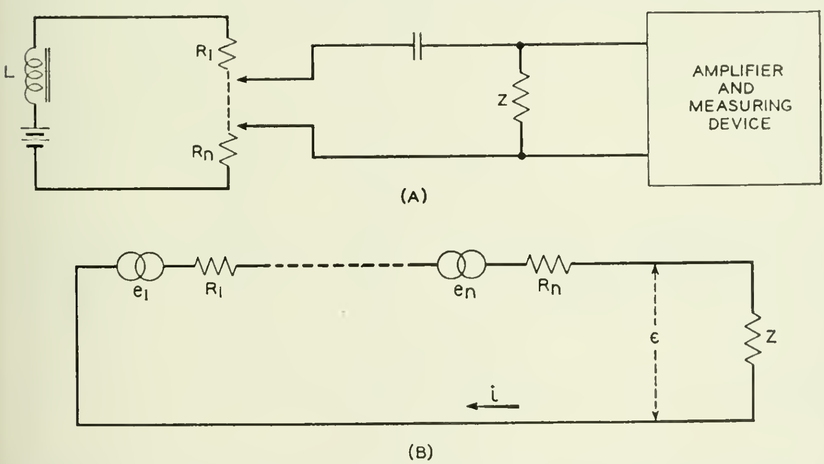


Fig. 13—(*A*) Circuit for measuring the contact noise of a series assemblage of resistance elements. (*B*) The equivalent a.-c. circuit of (*A*).

a.-c. circuit is shown in Fig. 13*B*. If the value of $\overline{\epsilon^2}$ impressed on the impedance Z of a measuring device is calculated, in a way similar to that outlined for the case of contacts in parallel, one gets

$$\overline{\epsilon^2}_{\text{series}} = \frac{[\overline{e_1^2} + \cdots \overline{e_n^2}]Z^2}{[R_1 + R_2 + \cdots R_n + Z]^2}. \tag{7}$$

Equations (6) and (7) are derived by considering the aggregate as made up of single contacts, but the same would be true if the aggregate were considered as made up of unit cells of arbitrary size, the values of e and R applying to the unit cells.

It is experimentally impracticable, if not impossible, to determine the noise and resistance of each contact in an aggregate; hence in the experimental test of the foregoing equations we have, in effect, divided the aggregate into unit cells. The aggregate of carbon granules was placed in four parallel grooves cut in a single phenol fibre block. Each groove was provided with five gold electrodes evenly spaced along the surface of the groove so that there were, effectively, four contiguous carbon cells in each groove. This permitted the measurement of the noise from each cell separately and also from the cells in various assemblages. Applying Eq. (6) or (7), as the case demanded, to the values of the noise from the single cells we calculated the expected noise of the assemblages and compared this with the measured values. Table I gives typical results for parallel assemblages and Table II typical results for series assemblages.

TABLE I

Cell	Cell Resistance Ohms	Noise in db	
		Measured	Calculated
a	3523	15.7	
b	4315	17.2	
c	4515	19.6	
d	3585	15.2	
$a + d$ parallel	1760	12.6	12.5
$b + c$ parallel	2215	15.7	15.5
$a + b + c + d$ parallel	980	11.2	10.9

TABLE II

Cell	Cell Resistance	Noise in db	
		Measured	Calculated
1	590	31.5	
2	720	34.0	
3	610	33.0	
4	510	30.0	
5	490	30.5	
6	545	31.0	
7	685	33.5	
8	427	26.5	
9	405	25.0	
10	226	19.0	
11	152	16.5	
Series 1 to 4	2510	38.5	38.2
5 to 8	2211	37.5	36.9
9 to 11	775	26.0	26.4
1 to 8	4720	41.0	40.4
5 to 11	2980	37.5	37.2

Considering the difficulty in holding a fixed granular configuration in a cell of loose contacts we feel that the experimental results justify the conclusion that the assumptions underlying the derivation of Eqs. (6) and (7) are essentially correct, which require that the phase of the noise voltage from each unit cell is entirely independent of that of any other cell in the aggregate. From this we conclude that the mechanism causing the noise is a small-scale effect capable of independently existing within a volume element much smaller than the size of the unit cell in any of the experiments we have performed on aggregates. In fact, as will be assumed later, we believe the noise mechanism to be located in a volume element smaller than that concerned with the properties of a contact between two particles.

If resistance elements are so chosen that each has the same resistance and noise, and these are placed in a circuit where the impedance Z is large compared to the resistance of the elements, then Eqs. (6) and (7) can be written respectively as follows:

$$\overline{V_c^2}_{\text{parallel}} = \frac{\overline{e^2}}{n}, \tag{6a}$$

and

$$\overline{V_c^2}_{\text{series}} = n\overline{e^2}. \tag{7a}$$

The resistance R of a parallel assemblage of like contact elements, each having the same resistance R_k, is obtained from $1/R = n/R_k$, or $n = R_k/R$.

Substituting this in Eq. (6a) we get

$$\overline{V_c^2}_{\text{parallel}} = \frac{R\overline{e^2}}{R_k}. \tag{6b}$$

For like contact elements in series we get $n = R/R_k$, hence Eq. (7a) becomes

$$\overline{V_c^2}_{\text{series}} = \frac{R\overline{e^2}}{R_k}. \tag{7b}$$

If now we have the further condition that the battery voltage is so adjusted for each new assemblage that $\overline{e^2}$ is always constant, then Eqs. (6b) and (7b) are equivalent and we have

$$\overline{V_c^2} = \text{Const. } R. \tag{8}$$

An equivalent relationship was derived and experimentally tested by Otto,[5] but it is clear from our derivation and measurements that it applies only to a change in the assemblage of like contacts, such as is

realized when the dimensions of a granular aggregate or conducting film are altered, and is not valid for cases where the resistance is altered by changing the contact compressions. In this latter case Eq. (2) applies.

Another interesting property of an assemblage is obtained if we express $\overline{e^2}$ by means of Eq. (1) in terms of the battery voltage V. Thus for the parallel assemblage, $\overline{e^2}$ = Const. V^α, and for the series assemblage, $\overline{e^2}$ = Const. $\left(\frac{V}{n}\right)^\alpha$. Thus Eqs. (6a) and (7a) can be written, respectively, as follows:

$$\overline{V_c^2}_{\text{parallel}} = \frac{\text{Const. } V^\alpha}{n} \tag{6c}$$

and

$$\overline{V_c^2}_{\text{series}} = \frac{\text{Const. } V^\alpha}{n^{\alpha-1}}. \tag{7c}$$

If we now accept as an approximation $\alpha = 2$ then Eqs. (6c) and (7c) are equivalent, and we can say that for any assemblage of contacts, where the value of $\overline{e^2}$ for each contact element is equal to that of every other contact element in the assemblage, the contact noise of the assemblage is inversely proportional to the number of contact elements in the assemblage. This principle we have established experimentally by building "square" assemblages—$\sqrt{n}$ parallel paths with $\sqrt{n}$ elements in series in each path—and measuring the noise as a function of n. The "square" assemblage is particularly interesting for it allows a control of the noise of an assemblage without altering its overall resistance. This suggests a principle which may be followed in designing grid leaks and carbon transmitters with low contact noise characteristics.

Discussion

It seems to us that the most logical hypothesis consistent with the foregoing experimental data is, as before indicated, that the noise mechanism lies in a fluctuating contact or boundary resistance. Assuming this we are led to the following considerations concerning the nature of the noise mechanism.

Careful measurement has established that the conduction through a carbon contact, as near as can be observed, is entirely ohmic. We have shown that when a carbon contact through which direct current is flowing is cyclically compressed, as in the acoustic modulation of a carbon transmitter, the generated a.-c. power is proportional to the square of the d.-c. voltage. This leads to the conclusion that the

resistance modulation due to the cyclical compression is independent of the applied voltage. It is evident from Eq. (1) that the fluctuating resistance responsible for the noise cannot be equivalent to the resistance modulation introduced by a cyclical compression where the contacting granules move relatively as a whole, for the fluctuating resistance responsible for noise is somewhat voltage sensitive as indicated by the departure of α from the value 2. This means either that the conductance responsible for the noise is specifically non-ohmic or that the extent of the conduction wherein the noise mechanism lies is diminished as the applied d.-c. voltage is increased. Non-ohmic conductance is usually such that conductance increases with applied voltage, thereby demanding a value of α in Eq. (1) which is greater than 2; accordingly we are inclined to believe that applied voltage acts to diminish the area over which the noise mechanism operates.

If the noise mechanism were intimately associated with the total conductance of a contact one would expect the noise to be proportional to some simple integral power of the current in a contact, but this is denied by the observed value of β in Eq. (2).

These facts and deductions lead us to the hypothesis that there exist two mechanisms of conduction between particles in contact, a primary conduction which accounts for the major portion of the current, and a secondary conduction wherein a relatively small portion of the total current is transferred and in which the noise mechanism is found. Goucher [1] has given evidence that the primary conduction between contacting carbon particles is of the same nature as that in solid carbon, and since we have been unable to measure any noise in solid carbon we assume that the secondary conduction does not take place through the same region of the contact as the primary conduction.

Recent investigation of the elastic nature of carbon contacts [1] has led to the conclusion that the surface of each particle can be considered as covered with a layer of hemispherical hills of heights distributed according to the function $N_x = \text{Const. } x^n$, where N_x is defined as the quantity which when multiplied by dx gives the number of hills coming into coincidence with a plane as it moves from the position x to $x + dx$, and n is a constant whose experimentally determined value is about 0.6. The establishment of a contact consists in bringing into coincidence a number of these hills and enlarging the coincidence areas to the extent demanded by the displacement of the contacting elements after their initial coincidences. Let us accept this picture of a contact and inquire as to how it applies in the explanation of our empirical noise equation.

We assume that through each area of coincidence the primary con-

duction takes place and that secondary conduction, in which the noise mechanism lies, can take place between the surfaces which are not in primary contact and are not separated by more than a certain increment dx. Figure 14 is an attempt to picture a portion of the hypothetical plane of contact between two carbon granules.

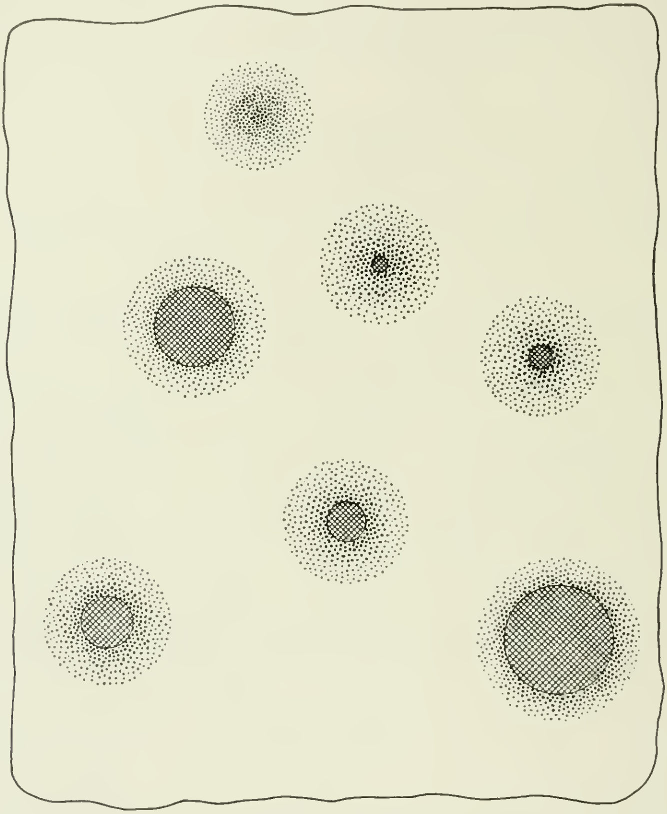


Fig. 14—A portion of the hypothetical plane of contact between two carbon granules. The cross hatched circular areas are the coincidence areas through which primary conduction takes place. The shaded areas, through which the secondary electrical conduction takes place, are regions where the granule surfaces are not separated by an interval larger than Δx. The area of each of these is independent of the coincidence area it surrounds.

The total number of hills in coincidence in a contact is given by

$$N_c = \int_0^D N_x \, dx = \text{Const.} \int_0^D x^n \, dx = \text{Const.} \, D^{n+1},$$

where D is the total displacement from the first coincidence. This number can also be expressed in terms of the contact resistance R by using Goucher's [1] derived equation: $1/R = \text{Const. } D^{n+3/2}$. Thus the above expression can be written

$$N_c = \text{Const. } R^{-(2n+2)/(2n+3)}. \tag{9}$$

The area of secondary electrical conduction surrounding each area of coincidence is precisely that area which would be added were the contact compressed by an increment of compression Δx. From the theory of Hertz [10] one can show that for smooth spherical hills in contact, $\Delta A/\Delta x$ is independent of the total hill compression and hence that the total area of secondary electrical conduction in a contact A_c is proportional to N_c, giving

$$A_c = \text{Const. } N_c = \text{Const. } R^{-(2n+2)/(2n+3)}. \tag{9a}$$

For purposes of analysis let us think of the secondary conduction area surrounding each primary area of contact as divided up into small elements of like nature, and further that the secondary conduction through each of these elements of area is independent of that in every other element. If such a contact is connected in a circuit as shown in Fig. 12A then the equivalent a.-c. circuit can be thought of as that shown in Fig. 15, where R is the mean resistance of the entire contact,

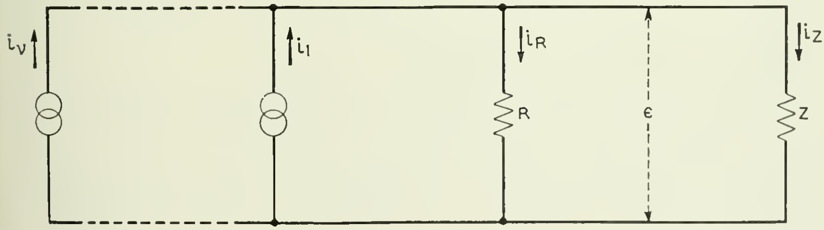


Fig. 15—The equivalent a.-c. circuit of a contact with ν elements of secondary conduction area, through each of which there is an independently fluctuating current, when such a contact is connected in a circuit as shown in Fig. 12A.

$i_1, \cdots i_2$ are the instantaneous deviations of the current from the mean current in each of the ν elements of secondary conduction area, and ϵ is the instantaneous value of the fluctuation voltage across the measuring device with impedance Z. Taking into account the random nature of the fluctuation currents through each element of secondary

[10] A. E. H. Love, "Mathematical Theory of Elasticity," 2nd ed., p. 192. This has been experimentally confirmed by J. P. Andrews, *Phys. Soc. Proc.* **43**, 1 (1931).

conduction area one can derive, in much the same manner as for Eq. (6), the relationship

$$\overline{\epsilon^2} = \nu \cdot \overline{i^2} \cdot (RZ)^2/(R + Z)^2, \tag{10}$$

where $\overline{i^2} = \overline{i_1^2} = \cdots \overline{i_\nu^2}$. Since ν is proportional to A_c we have by Eq. (9a) the relationship $\nu = \text{Const.}\ R^{-(2n+2)/(2n+3)}$. Substituting this into Eq. (10) we get

$$\overline{\epsilon^2} = \text{Const.} \cdot R^{(2n+4)/(2n+3)}\, Z^2/(R + Z)^2,$$

and if we make Z large compared with R this becomes

$$\overline{\epsilon^2} = \text{Const.}\ R^{(2n+4)/(2n+3)}. \tag{11}$$

Comparing this with Eq. (2), the experimentally derived relationship between noise and contact resistance, we get $n = 0.5$. Goucher [1] found by elastic measurements the value of $n = 0.6$.[11] In view of the fact that a slight change in the experimental value of the exponent β in Eq. (2) causes a rather large change in the value of n thereby determined from Eq. (11), we can say that the agreement between the results obtained from elastic measurements and noise measurements is surprisingly good, and that this agreement supports the hypothesis regarding the nature of a contact.[12]

In discussing the hypothesis that there exists a region of secondary conduction which is responsible for the noise, we have not made any assumption as to the nature of the secondary conduction. Several possibilities, however, have occurred to us, one of which it seems desirable to mention at this time.

If one assumes that the thermo-mechanical vibrations of a solid extend to the outside surface, then it is possible that the wave crests may be able to make periodic electrical contact across the secondary conduction area assumed in our hypothesis. This would permit a pulsating current to flow, the frequency of which, it is supposed, is determined by the frequency of the oscillator. For oscillators of audible frequencies the law of energy equipartition applies and each oscillator will have the usual $1/2\ kT$ of energy per degree of freedom. The energy of an elastic oscillator is also proportional to $(B \cdot F)^2$, where

[11] Goucher found a discrepancy between the measured resistance-displacement, resistance-force, and force-displacement relationships. But for reasons stated in his paper we are inclined to accept the distribution function found from the force-displacement measurements.

[12] If one assumes the existence of a film in the contact, which some older theories of microphonic action do, and that the number of independent elements of area through which current is conducted is proportional to the area of this film then one is led to the very unsatisfactory conclusion that $n = -3.5$.

B is the amplitude and F the frequency. From these two expressions of the energy we get $B \sim F^{-1} T^{1/2}$. The area of secondary electrical conduction surrounding each hill in coincidence is proportional to B, and since this area determines the number of elements of area through which secondary conduction takes place, as assumed in the derivation of Eq. (11), we arrive at the conclusion that

$$\overline{\epsilon^2} = \text{Const. } F^{-1} T^{1/2}. \quad (12)$$

While the hypothesis leading to Eq. (12) is only intended as a suggestion it does explain the inverse frequency relationship, and the temperature relationship is not an impossible one judging from the past unsatisfactory measurements. It may be possible, also, to explain the departure of α in Eq. (1) from the value 2, for one would expect electrostatic forces to distort the contacting surfaces so that the secondary conduction area would become smaller as the voltage increases. A satisfactory experiment on the effect of temperature on noise will do much to establish or disprove the tenability of this hypothesis.

Brillouin [13] has recently derived an expression for the noise in a conductor carrying a current by using the statistical method to deduce the most probable distribution of the electrons in such a system when it is in equilibrium. This method of calculation gives a noise energy, in addition to thermal noise, which is proportional to the square of the current and inversely proportional to the volume of the conducting material. We have made a calculation of the relative magnitudes of the two terms in Brillouin's equation which correspond to our experimental conditions. Assuming reasonable dimensions for a carbon contact and a current density as high as any we used it turns out that the magnitude of the term for contact noise is far below that for thermal noise. Furthermore it seems to us that Brillouin's mechanism would require a flat frequency distribution of noise rather than the distribution which we have observed. For these reasons we do not believe that the noise which we have studied is produced by the mechanism postulated by Brillouin.

In conclusion we wish to acknowledge our indebtedness to Dr. J. B. Johnson and Dr. F. S. Goucher for the helpful criticism they have given us during the course of this work.

[13] L. Brillouin, *Helv. Phys. Acta, Supt.* 2, **7**, 47 (1934). This theory is intended to explain the "Fluctuations de résistance dans un conducteur métallique de faible volume," reported by M. J. Bernamont, Comptes Rendus **198**, 1755 (1934); ibid. **198**, 2144 (1934).

Contemporary Advances in Physics, XXX—The Theory of Magnetism *

By KARL K. DARROW

The topic of this article is the explanation of magnetism as ordinarily observed—to wit, the magnetization of pieces of matter of ordinary dimensions—by ascribing magnetic moment to the individual molecules, atoms, and electrons of which matter is composed. For *paramagnetic* bodies it is postulated that the individual atoms are magnets of which the orientation, but not the strength, is altered in the presence of a magnetic field; the theory is so successful as to make it possible to calculate, from magnetization-curves, values for the magnetic moments of these atoms which agree admirably with those deduced from spectroscopic theory and from experiments of other types. For *ferromagnetic* bodies the same postulate is made, but it is necessary in addition to recognize the existence of huge interatomic forces of which very little is known, so that a large proportion of the science of ferromagnetism still lies beyond the scope of atomic theory. For *diamagnetic* bodies the phenomena are interpreted in a simple and effective manner, as an immediate corollary of the well-known structure of the atom.

MAGNETISM is a quality which we attribute to the atom. We affirm that iron, nickel, gadolinium, gaseous oxygen, and in fact all substances, are magnetic because there is magnetism in their atoms. Indeed we go even deeper, and affirm that the individual electrons and the nuclei within the atoms are magnetic. Nevertheless, the atomic theory of magnetism is a really valuable theory. Perhaps that "nevertheless" sounds out of place; but I assure you that without it there would be a trace of paradox in the statement, which perhaps our grandfathers would have been quicker at discerning than are we. Let me explain my meaning by referring to the atomic theory, or as it is usually called the kinetic theory, of gases. Those who designed this theory succeeded in explaining the pressure, the temperature, and the viscosity of gases, without attributing a single one of those qualities to the atoms. To the atoms they assigned the properties of momentum and velocity and kinetic energy; those other qualities which I just named were then interpreted in terms of these,—they were interpreted as what we call *statistical* properties of the great multitude of atoms which constitutes a gas. This was a real explanation of pressure and viscosity and temperature, in the fullest sense of the word "explanation"—or anyhow, in the fullest sense of that word which is customary in physics. But along with these properties of pressure and viscosity

* Expanded from a lecture delivered on January 14, 1936, at the School of Engineering of Yale University, and still bearing obvious traces of its original form. In preparing it I received invaluable aid from Dr. R. M. Bozorth.

and temperature, a gas also possesses weight. The builders of the kinetic theory simply said that the weight is a property of the individual atoms, and that the weight of the gas is the sum of the weights of its atoms. Now evidently this was not an explanation of weight at all. Indeed, by assigning weight to the individual atom, the builders of the theory had foregone all attempts at an explanation. A property which you assign to the atom is a property which you refuse to try to explain in terms of the atom—or so at least it always seemed to our forefathers. To assign a quality to an atom used to be taken as a confession of incompetence to explain that quality. I can of course make this clear by proceeding to absurd extremes. If I say that an orange consists of soft yellow juicy atoms, or that a marshmallow is made of sweet white sticky atoms, or that a piece of iron is made of hard black shiny conductive atoms, you recognize at once that those are not serious atomic theories: they are just futile and somewhat ridiculous statements. If I claim to explain the weight of a piece of iron by saying that it is the sum of the weights of the atoms, I am making a claim which unfortunately may not sound ridiculous, but is really just as futile—unless it acquires value by being linked with some other assertion. But when I say that the magnetism of a piece of iron is due to the magnetism of its atoms and its electrons, the statement is by no means a futile one; it is significant and important. For this there are two main reasons or rather groups of reasons, which I will indicate by the words *orientation* and *atomic structure.* (In addition there are remarkable experiments on jets of atoms whereby their magnetic moments are measured directly, but these I reserve for another occasion.)

First a few words about atomic structure. It is a fact of experience—the experience of one hundred and fifteen years—that a current running around a loop of wire is the equivalent of a magnet. If now somebody asserts first that a piece of iron is magnetic because its atoms are magnets, and then goes right ahead and asserts that the atoms are magnets because they have perpetual currents running around inside them—well, the combination of these two statements is not necessarily futile or trivial. At the very least, it is a sensible attempt to reduce the two kinds of magnetism apparently existing in the world to a single kind, that which is due to moving electricity. This was Ampere's idea a hundred years ago. Now if in addition there is independent evidence that the atom comprises mobile electrical particles, then this idea of Ampere's becomes the assertion that those particles inside the atom are actually revolving. It is well known that modern physics is full of such evidence, of evidence that atoms contain very mobile electrons; and some of my readers may recall that thirty years or so ago there

were an atom-model with stationary electrons and an atom-model with revolving electrons, which were in competition with each other, and the latter of which has by now driven the former utterly out of the field. Remember now, that the atom-model with the revolving electrons triumphed over the other one not primarily because of its magnetic quality, but because of the theory of spectra which Bohr and others were able to derive from it. Revolving electrons in atoms were first of all proved to be responsible for spectra, and then it was noticed that they are capable of causing magnetism. Therefore when the physicist says that magnetism is a quality of atoms, he is not making a confession of incompetence, but an inference from a highly-developed and successful theory of spectra; and this makes all the difference in the world to the value of the statement. Indeed the situation is even better than I have intimated; for there are dozens of cases in which first an atom-model or a molecule-model has been constructed expressly to explain the spectrum of the substance in question—then, the magnetic moment of its system of revolving electrons has been computed—then, the magnetic moment of the atom or the molecule has been measured—and the two have agreed! This is really an understatement, which needs to be broadened so as to include the cases in which the spin of the electron plays a part; but I pass them over, intending to defer the broadening to the latter part of the article, which is to be devoted to these matters of atomic structure. For the moment, let me make just one more allusion to them, a very important one. Electrons revolving in orbits around a nucleus obviously possess angular momentum. Therefore, if an atom has a magnetic moment due to revolving electrons, it has an angular momentum also. This again is an understatement, for it contains a restriction which can be removed in view of the broadening which is later to be made. It appears to be a general rule that *in the atom, magnetic moment and angular momentum always go together.* A magnetic atom is a gyroscope—necessarily and automatically. This is a fundamental principle, and from it flow some strange and striking consequences. Everyone who has worked or played with the classic gyroscope of our laboratories knows that it has quaint and tricky idiosyncrasies. Well! the atom has them too; but it has others in addition. Angular momentum, on the atomic scale, is subject to peculiar laws of quantum mechanics; and the atomic magnet-gyroscope behaves in extraordinary ways, of which our laboratory gyroscopes give not the faintest intimation.

To summarize my introduction then: the first step in the theory of magnetism consists in referring it to the individual atom. This sounds like a confession of defeat, but it is nothing of the sort; it is a claim of

victory. Our theory of spectra requires that atoms, or some of them at any rate, should be magnets, and they *are* magnets. Moreover it fixes the magnetic moments which certain atoms ought to have, and so far as our experiments go, the atoms *do* have these moments. Moreover it imposes angular momentum on these atoms, and fantastic as the consequences are, experience bears them out. Logically, then, I should begin the main part of my talk by showing how the magnetic moments and the angular momenta of atoms and of molecules are calculated from their spectra by atomic theory. This, however, would by itself require several lectures, and very difficult ones at that.* I must therefore simply ask you to believe that the magnetic moments of atoms are inferences from fundamental theory, not mere *ad hoc* assumptions; and now I will explain what I had in mind when I wrote down the word *orientation* to designate one of the topics of this article.

It is one of the best-known facts of physics that the magnetization of a substance is not fixed and constant, but increases with the strength of the magnetic field which is acting on the substance. By the way, before going any further I must definitely exclude the so-called "diamagnetic" substances. That exclusion being made, we do *not* assume that the magnetic moment of the individual atom increases similarly with the field strength. People did not make that assumption, even in the days before the fundamental theory was developed. Had they done so, it would have been just as silly as saying that a marshmallow is made of soft white sticky atoms, and calling *that* an atomic theory. They supposed, and we suppose, that the moments of the individual atoms remain practically the same whatever the field strength; what changes is the average *inclination* of these moments to the field. The atomic moments are vector quantities pointing in various directions, different from one atom to the next. The magnetization of the substance as a whole is the resultant of all these myriads of tiny vectors pointing in their various directions. If they all pointed the same way the substance would be completely and perfectly magnetized, with a moment equal to the total number of the atoms multiplied into the moment of any one. This state of *saturation* is not, however, to be attained, not even to be approached without a rare and felicitous concourse of a favorable substance, a very low temperature and a very strong field. Much easier of attainment is the opposite extreme, when the vectors are pointing all ways at random and the magnetization is zero. This happens with nearly all substances when there is no field applied, and it seems quite natural. But when even the smallest field

* This subject was partially treated in "Contemporary Advances in Physics, XXIX . . . ," April 1935 *Bell Sys. Tech. Jour.*

strength is applied to such a substance, you might expect all the little magnets to turn right around and point straight up the field-direction, achieving saturation in an instant. Well, it is certain that saturation is not achieved; but still there is some degree of magnetization, as though the little magnets all started to turn around and were stopped before they got very far. What is it that might stop them? If you look at the books of twenty or twenty-five years, you will find an answer: they are stopped by the collisions which these atoms make with one another. This is the classical idea, which is generally thought to be well verified by experiment. But let us look into the matter a little more closely.

For simplicity let us imagine a gas—preferably, unit volume of the gas—composed of N identical atoms, each with a magnetic moment μ, and exposed to an applied field H. Visualize some particular atom, of which the career is an endless alternation between free flights and sudden impacts. All the time the magnetic moment of the atom, the little vector of which I have been speaking, is subject to a torque arising from the field. The classical idea is, that throughout every free flight that torque is steadily bringing the vector more and more nearly into alignment with the field, but usually not having time enough to succeed, because at every collision the vector is suddenly and violently redirected in a perfectly arbitrary way. Gradual approach to alignment during the free flights, violent dis-alignment at the collisions, and the magnetization of the substance indicating how far the alignment progresses, on the average, before the dis-alignment stops it—this is the classical picture. It all seems beautifully obvious, and yet is it now believed to be entirely false!

The trouble lies in the fact that the atom is a gyroscope. You recall that it is one of the oddities of the gyroscope that when you apply a torque to it, it starts off at right angles to the direction in which you expect it to go. Now here is our atom just leaving the scene of a collision with its magnetic moment making, say, an angle ϕ with the field-direction. As it flies away the torque is steadily trying to reduce the value of ϕ, but instead of obeying, the atom just keeps on blandly precessing about the field-direction, the value of ϕ remaining obstinately the same. The unbreakable link between magnetic moment and angular momentum has neatly killed the supposition that the field magnetizes the gas because it aligns the atoms, or partially aligns them, during their intervals of unimpeded flight. The free flights are just the periods when nothing whatever happens in the way of alignment. Much labor has been expended in the hope of finding some way out of this impasse, but none has been revealed except that of supposing that

whatever is the mechanism whereby the field aligns the atoms, it is a mechanism which operates during the impacts and not between them. Partial alignment *at* the collisions, no change in the situation during the free flights—this amounts pretty nearly to standing the classical theory on its head!

Nevertheless the mathematics of the classical theory remains entirely unchanged. This is because the mathematics merely expresses the assumption that the field has managed to find *some* way of partially aligning the atoms, and does not concern itself in the least with what that way may be. This sounds rather vague, so let me remind you just what the assumption is. Suppose to begin with that the vectors of the atoms are capable of only two orientations in the applied field: one parallel, the other anti-parallel to the field-direction. To transfer an atom from the one orientation to the other, we must do work against the torque of the field (or receive work from the torque of the field) amounting to $2\mu H$. We have, therefore, two classes of atoms, differing in energy by $2\mu H$. Let N_1 and N_2 stand for the numbers in these classes at some particular instant. Now the classical theory, as I have been calling it, is strictly no more than the assumption that the ratio of N_1 and N_2 is given by Boltzmann's theorem:

$$N_1/N_2 = \exp(-2\mu H/kT) \tag{1}$$

and the essence of this assumption, I take it, is that the atoms are able to change their orientation so as to pass from either class to the other, and that they employ this facility of free passage to get themselves into thermodynamic equilibrium at the temperature T of the gas. This has been the assumption ever since Langevin founded the theory, and it still is the assumption, even though we may no longer enjoy that pretty picture of the mechanism of the change of orientation which once we accepted, and have no other to replace it.

I can readily write down the complete theory of this case. We introduce the two additional equations,

$$N_1 + N_2 = N, \qquad I = (N_1 - N_2)\mu, \tag{2, 3}$$

of which the first says that all the atoms are in either the parallel or the anti-parallel class, and the second that the magnetization I of the unit volume of gas is the resultant of the vectors of all its atoms. Now we eliminate N_1 and N_2 between the three equations, and swiftly arrive at the result:

$$I = N\mu \tanh(\mu H/kT), \tag{4}$$

which is the equation of a curve starting obliquely off to the right from

the origin, and bending over to approach an asymptote which is a horizontal line of the ordinate $N\mu$.

At this point a strictly classical physicist would certainly grin or sneer, because he would say to himself: "The speaker started out by assuming for simplicity that the atoms can point in only two directions, and now he has gone on to his conclusions without remembering the obvious fact that an atom may point in any direction whatever!" Well, of course Langevin did make allowance for that supposedly obvious fact; it complicates the affair to some extent, but not seriously, and leads to a very similar curve for I versus $\mu H/kT$. Quantum mechanics, however, flatly denies that it is a fact. I mentioned above that the atomic gyroscope has some paradoxical properties of its own, in addition to those which it shares with the laboratory gyroscope. Here is one of them. *The atomic magnet is supposed to be able to set itself*, not at any angle whatever with respect to the applied magnetic field, but *only at one or another of a small number of definite discrete angles*. This is because of its angular momentum: it is primarily the angular momentum which is constrained to this very singular behavior, and which the magnetic moment is automatically obliged to follow because they are so closely linked together. If I am asked why the angular momentum should behave like this, I can only reply that according to what I am told, if one is sufficiently penetrated with the spirit of quantum mechanics it seems self-evident, and if one is not sufficiently penetrated with that spirit there is nothing which can be done to help. Notice anyhow that it is compatible with the statement that when the atom is freely flying along, the field just keeps it precessing about the field-direction, instead of gradually aligning it; and there is ground for being thankful that this derivation, and certain others, are somewhat simplified by it. It may be asked, how many different inclinations are permitted to the atom? This depends upon the angular momentum of the atom, and we can tell it from the spectrum. There are certain elements and certain compounds for which the case is just as simple as I have described it; just two permitted orientations, the parallel and the anti-parallel, and no more. There are others for which the permitted inclinations are three in number, others for which there are four, five, and other integers up to fifteen or twenty. All these yield curves of I versus $\mu H/kT$ having the same general traits, but differing in the rate at which they approach the asymptote. I will refer to all such curves as "Langevin curves," although the only one which Langevin proposed was the classical curve corresponding to the case in which all orientations are permitted (or, as we may say, there are infinitely many permitted orientations).

You may now be expecting me to say that there are many gases, and possibly other substances as well, for which experimental curves have been obtained that are comparable with these. I am obliged to disappoint you. You can readily see that in order to get over onto the "curvy" part of these curves, one must work in experimental conditions in which the argument $\mu H/kT$ is greater than, or anyhow not very much less than, unity. One thinks, of course, of using the highest accessible field strengths H so as to enhance the numerator of that fraction. This, however, is not sufficient, for it turns out that μ (the magnetic moment of an atom or a molecule) is so very small that one is obliged to diminish the denominator also by going to the lowest attainable temperatures. All the experimental curves of this character have been obtained at temperatures lower than 15° absolute, some at temperatures between 1° and 2° absolute. This excludes all the gases. Moreover, it has been necessary thus far to choose the atoms with the largest magnetic moments, and these turn out to be, quaintly and inconveniently enough, the atoms of the rare-earth elements. Probably the best of the experimental curves (Fig. 1) relates to a substance which most people never have heard of, in this or any other connection: it is gadolinium sulphate. There are about a score of such curves ob-

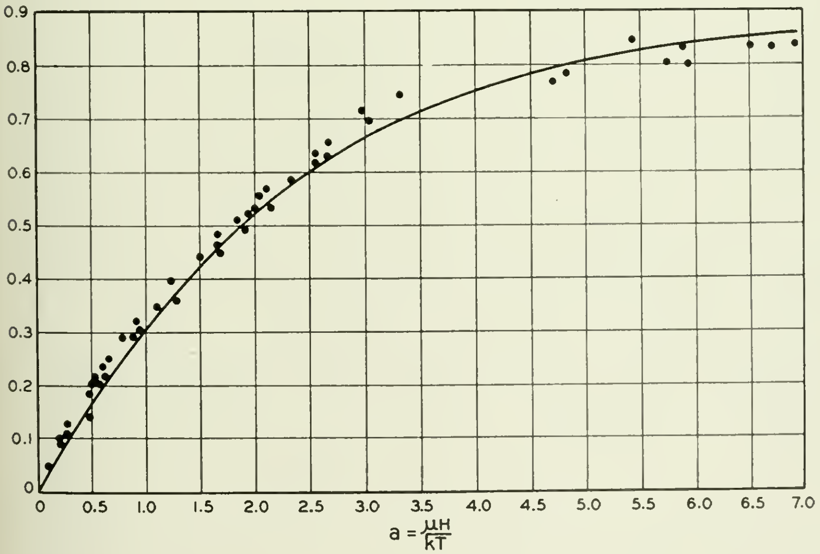


Fig. 1—Magnetization of a paramagnetic salt [$Gd_2(SO_4)_3 + 8H_2O$] as function of the parameter a; the ordinate is I referred to its saturation-value (deduced by extrapolation) as unity. Data from Onnes and Woltjer. The curve is the "classical" Langevin curve (number of permitted orientations, $n = \infty$) which is hardly distinguishable from the quantum-theory curve for this particular case ($n = 15$).

tained with minerals and glasses containing these elements, and most of them agree well with one or another of the theoretical curves; in which connection there is an interesting detail, which I will bring up at the end of the article. One would scarcely expect a theory worked out for gases to apply so well to solids, and as a matter of fact it is a peculiarity of the rare-earth atoms that even when incorporated in a compound or a solid they behave in several ways more like the atoms of a gas, than happens with any other elements.

Pray do not think, however, that all this time I have been talking about a theory which has no application excepting to the rarest of all elements under the rarest of all temperatures. Its applications are a good deal wider than that. True it is that with gases universally, and with other substances ordinarily, we cannot get data along the curvy parts of the curves; but we can make measurements along the sensibly-linear parts near the origin. This amounts to saying that we can determine the slope of the curve at the origin. Now of course it sounds ridiculous to speak of confirming a theoretical curve by measuring its tangent at one point. In this case, however, it is not altogether ridiculous. Usually the experiments are made by varying H and measuring I while the temperature is kept constant. Suppose this is done for several different temperatures, and suppose the results are plotted by using H instead of $\mu H/kT$ for the abscissa. Then the theory supplies us with different curves for the different temperatures, all having the same general aspect, but different slopes at the origin. I will denote these slopes for the time being by $tan\ \theta_0$. The theory, then, requires that $tan\ \theta_0$ should be proportional to $1/T$; and for gases, this is found to be the case. Of course this is not such good evidence for the theory as would be a complete following-up of the curve nearly all the way to the asymptote; but it is pretty good by itself, and for further evidence we can invoke those experimental curves for gadolinium sulphate and other solids of which I just spoke.

If now we let ourselves be convinced by this evidence, a valuable conclusion follows. From the slopes of these curves at the origin, the value of μ can be deduced. Let us go back to the curve of I versus $\mu H/kT$ or a, which is the epitome of all the rest. We write:

$$dI/da = N\mu(1 - \tanh^2 a), \tag{5}$$

$$(dI/dH)_{T=\text{const.}} = (dI/da)(da/dH) = (1 - \tanh^2 a)N\mu\cdot(\mu/kT). \tag{6}$$

Since measurements are actually made at a fixed temperature and refer to the slope of the curve near zero field strength, we evaluate this derivative for $a = 0$, and we get:

$$\tan\theta_0 = (dI/dH)_{H=0} = N\mu^2/kT, \tag{7}$$

and thus from the measurement of θ_0 at any temperature we derive the magnetic moment of the individual atom or molecule of the gas. This formula ought to give the right order of magnitude for μ in any case. Whether or not it will give exactly the right value, will depend on the validity of one of the assumptions, which I now recall. This particular formula is for the case in which the atoms have only two permitted orientations in the field, the parallel one and the anti-parallel one. Had we supposed that every inclination is a permissible one, we should have arrived at $(1/3)N\mu^2/kT$. Had we supposed a number of permitted inclinations greater than two and less than infinity, we should have arrived at some intermediate value. So, I now write as the general formula,

$$\text{volume-susceptibility } \chi = bN\mu^2/kT, \qquad b = 1 \text{ to } 1/3, \tag{8}$$

having placed on the left the name and the symbol by which is generally known what I have been denoting by *tan* θ_0, and on the right a factor b of which the value will depend on the number—I will call it n—of permitted orientations, but will fortunately never be outside of the narrow range between unity and 0.33.

Thus a rough estimate of an atomic moment may be made without knowing the number of the permitted orientations. Very many such estimates have been made, and they always give values of μ quite compatible with what we know in general about the structures of the atoms. If we want to make an estimate truly accurate enough to serve as a stringent test of theory, then we must take from the spectrum of the atom, not only the spectroscopic value of magnetic moment with which we are going to make the comparison, but also the angular momentum of the atom which is what determines the number of orientations. This causes us no extra trouble, for if we understand the spectrum well enough to get the one we also understand it well enough to get the other. Now when we look into the literature to see how many such comparisons have been made, we suffer again a disappointment. It turns out that the noble gases and most other convenient gases exhibit the magnetic moment *zero*. This is of course no fortuitous bit of ill luck; it is the same thing, to wit a certain stable interlocking of the various electronic orbits and rotations in the atom, which leads on the one hand to a zero magnetic moment and on the other hand to a relative smallness of the forces which make for chemical combination and for condensation. Anyhow it is an inconvenience; but luckily there are two convenient gases, oxygen and nitric oxide—O_2 and NO—which do have magnetic moments different from zero; and the test of the theory is in these cases most satisfactory. The agreements be-

tween the magnetic moments calculated from magnetic data after this fashion, and those derived from the spectra, are accurate within an experimental uncertainty of a few promille. I think that these are among the most impressive results in the whole structure of modern physics. Then in addition the rare-earth elements help us out again, owing to the peculiarity which their atoms have of behaving, even when they are incorporated into solid compounds, as though they were the atoms of a gas. They have supplied us with a number of beautiful agreements of this same character.

Now as a transition to the next part of this paper, I must acquaint you with another fact which belongs to this last part. I have more or less been allowing you to suppose that with solids as with gases, the susceptibility is generally proportional to $1/T$. Actually it is much more common, among solids, to find a law of the type,

$$\chi = \text{const.}/(T - \theta), \tag{9}$$

where θ stands for a constant differing from one substance to another. This constant is evidently of the dimensions of temperature; it is a sort of "critical" temperature, known as the *paramagnetic Curie point;* the formula usually holds for a broad range of values of T above and not too close to θ. (There are plenty of cases where even this formula will not fit, but we will not concern ourselves with them.) You see that this might be taken as meaning, that for temperatures greater than θ the substance is more strongly magnetized by any particular field strength than, by our previous theory, we should expect it to be. It might even be taken as suggesting, that in addition to the applied field which we produce ourselves by a horseshoe magnet or something of the kind, there is an extra field arising within the substance itself, which helps along with the magnetization. Now this is just the suggestion which physicists have accepted. Of course it is necessary to make some specific assumption about this extra or internal field, in order to arrive at the empirical law which I just wrote down. The required assumption turns out to be simple and gratifying. It is necessary and sufficient to assume that inside the magnetized substance, there arises an extra field which is proportional to the magnetization I itself. Hitherto we have been supposing that the torque acting upon an atomic magnet is directly and entirely due to the applied field H, and we have been led to the law that χ varies inversely as T. Now we are going to suppose that the torque is due to a field $(H + AI)$; and this will lead us, by way of the equation

$$I = N\mu \tanh (H + AI)/kT \tag{10}$$

to the law that χ varies inversely as $(T - \theta)$. The constant A is one which we adjust so as to get the empirical value of the constant θ. It is pleasant to be able to say that this constant A —that is to say, the hypothetical extra field—does not meddle at all with the multiplying factor: the theory allows us to write:

$$\chi = \frac{bN\mu^2}{k(T - \theta)} \quad \begin{array}{l} \theta \text{ depending on } A \\ b \text{ depending on } n \end{array} \tag{11}$$

and (if it is the right theory, of course!) we can go on estimating atomic moments for substances of this category just as well as we can for the substances for which θ is zero. Most published values of μ correspond to such cases.

I am going to say very little about the extra field, or "Weiss field" as it is often called, because it is still one of the mysteries of physics. One realizes readily, of course, that if all the little atomic magnets turn themselves partially or totally into alignment, each one of them experiences a magnetic torque which is due to all the rest. It may be shown that this is proportional to the magnetization I, which looks very promising indeed; but alas, when it is calculated its magnitude turns out to be thousands of times too small. People used to say that AI must be a field of non-magnetic origin, which is just another way of saying the same thing. At present it is commonly believed that the force in question is what is called an "exchange" force, that is to say, an electrostatic force among electrons, of which the *modus operandi* can be discovered only by quantum mechanics. I am told that this quantum-mechanical theory has not yet been persuaded to deliver a really satisfactory result; but probably we shall be obliged to accept it in default of any other.

Now I call your attention to the fact that if the temperature should be made equal to or lower than θ, this last equation would predict something very wild and strange: an infinite, or a negative, susceptibility. This is a curious situation, and there are several cases in which we can appeal to experiment to resolve it. Take the elementary metal *nickel*, for example; if one measures the susceptibility over the range between 400° and 900° C. one gets a gorgeous curve of just this character, for which the value of θ is around 370°; now if one investigates nickel at temperatures below 370°, say around room-temperature, one learns that it is *ferromagnetic*. The same holds true for iron, for cobalt, for a diversity of alloys, except that θ varies from one case to another.*

* There are however cases in which the substance does not display the distinguishing marks of ferromagnetism (notably remanence) when $I < \theta$; and incidentally there are cases in which θ is negative; all of these are knotty problems for theory.

I will just refer to one additional case, because it is of very recent discovery and relates to that rare element which is so helpful in magnetics and seems to be so useless for anything else: I mean *gadolinium*. Metallic gadolinium has a value of θ amounting to about 300° absolute. Well, last spring Trombe at Strasbourg investigated this metal at low temperatures and found that it, too, is ferromagnetic, even more so than iron itself. Incidentally most of the rare-earth elements have not yet been prepared in pure metallic form, and it looks as though we might almost count on turning up some more cases of this kind. All this brings me to the question of *ferromagnetism*.

I do not suppose that any of my readers thinks that it is ferromagnetism of which I have thus far been speaking, but for the sake of completeness I will give the name: up to this point we have been considering *paramagnetic* bodies, and explaining their behavior by the orientations of atoms in fields. Now we turn to the properties of iron, cobalt, nickel, various alloys and compounds of these, various alloys containing manganese, and gadolinium: the *ferromagnetic* substances.

The most confusing thing about ferromagnetism—at least if my own experience as a student is any guide—the most confusing thing is, that the I-vs-H curve of a ferromagnetic substance reminds one of the sort of thing that the Langevin theory is meant to explain, and yet it is not that sort of thing at all. One looks at the Langevin curve with its approach to saturation, and then one thinks of the curve for iron with *its* approach to saturation, and one cannot help but think that the two must correspond to each other except for minor and trivial details. Well, they do *not*. They differ not alone in trivial details, but in every possible way, excepting the solitary common feature of the approach to a horizontal asymptote.

It is really impossible to put this statement too strongly. The Langevin curve and the iron curve differ in shape, as any sketch (cf. Figs. 1 and 2) will show. They differ utterly in scale. If I were to start to put a Langevin curve on the same plot where an iron curve appears with suitable detail, not only would it be sensibly linear for thousands and thousands of miles, but it would not even rise appreciably off the axis for hundreds of miles. Conversely if I had tried to put the curve for a ferromagnetic body upon the same graph as the Langevin curve, the former would have consisted only of the axis of ordinates plus the horizontal asymptote. Finally, the temperature relations are all wrong. I told you that in the Langevin curve the slope near the origin varies inversely as temperature, and I left you to infer that the ordinate at saturation is independent of temperature. In the curve for iron, the slope near the origin goes up with the temperature, and the

ordinate at saturation goes down as the temperature goes up. And yet, we interpret ferromagnetism by what is essentially an atomic theory: that is to say, we suppose that any piece of iron is an aggregate of little magnets each having a constant magnetic moment (so long as the temperature is kept constant) and that magnetization of iron consists in aligning these magnets.

I think it instructive to refer to these little magnets by the name of "atom," with some distinctive prefix; so, for a few minutes, I will call them "super-atoms," though this is not the customary name. When a piece of iron is unmagnetized or demagnetized, the super-atoms are

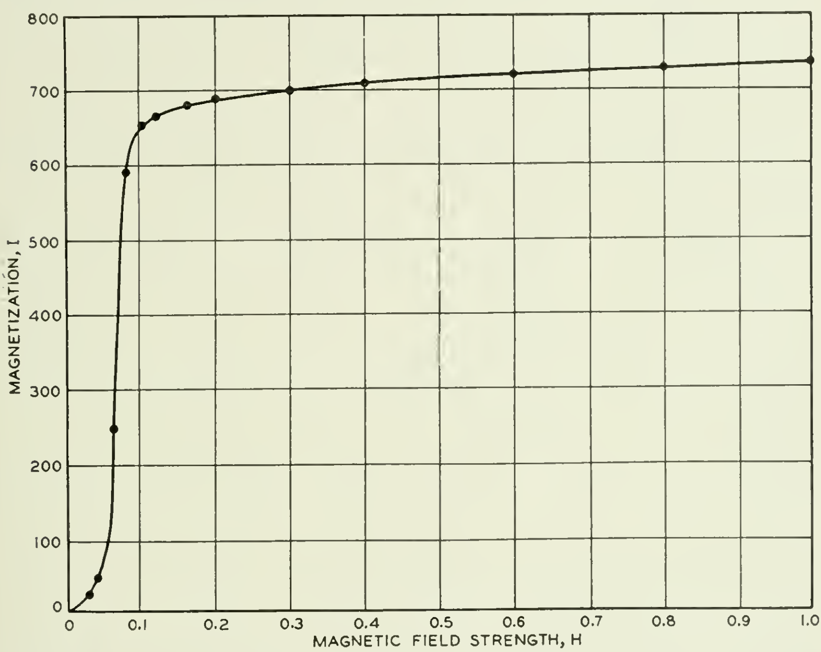


Fig. 2—Magnetization of a ferromagnetic material (81 permalloy, annealed two minutes at 1000° C.); the ordinate is I. Data by P. P. Cioffi.

pointing in all directions at random, just like the individual atoms of a paramagnetic gas which is unmagnetized. When a magnetic field is applied to the unmagnetized iron, the super-atoms get more or less aligned with one another. If the field is strong enough they are perfectly aligned, and there exists what is usually called "saturation" of the iron. Now it is worse than useless to remember about Boltzmann's theorem, or impacts, or free flights between impacts, for all those concepts have no relevance. We have to look at the phenomena, and see what they require.

We see at once that the super-atoms must be very easy to align, because saturation comes so quickly, with so relatively small a field strength. We learn also that when they are aligned, they are not exposed to the incessant urge to utter dis-alignment which afflicts the atoms of a paramagnetic substance, for iron continues to be magnetized when the field is withdrawn; not fully magnetized, as a rule, but considerably so. Heretofore I have been talking of substances, in which the atoms have a natural state of perfect dis-alignment or random orientation; a moderate field can derange it only a little, and the atoms return to it instantly and invincibly as soon as the field is cancelled. Now I am talking of substances in which the super-atoms seem to have no single natural state at all; a moderate field aligns them with ease, and when it is removed they like to linger in their alignment. The phenomena become clearer when we experiment not with ordinary iron, which is a chaotic mass of tiny crystals, but with a single large crystal. It turns out then that the super-atoms have a mighty preference for pointing along the cubic axes as distinguished from all the other directions; but as between these three cubic axes, and as between the two opposite senses along each of the three, they seem to be well satisfied with any. Suppose for definiteness that I have a cubic crystal of iron with one of its axes vertical, another in the meridian and the third, of course, pointing east and west. Then if the crystal is unmagnetized, one sixth of the super-atoms may be pointing east and one sixth west, one sixth pointing north and one sixth south, one sixth pointing up and one sixth down. (I do not say that this is necessarily the case, but it may be.) Now if I apply to the crystal a moderate magnetic field pointing north, the one sixth of the super-atoms which were already pointing north will not be affected, but all the other five sixths will flop right over and imitate them. It is amazing how small a field will suffice to do this: 100 cersteds for a good single crystal, whereas 100,000 oersteds, as I suggested, are not enough to bring the ordinary paramagnetic substance at room-temperature anywhere near to saturation. If next I cancel the field, the five sixths of the super-atoms which came over to the northward orientation will not be irrestibly urged to hasten back to their previous habit: indeed if I manage to avoid mechanical shocks and jarrings, most of them may linger indefinitely, still pointing in the direction to which the vanished field once tempted them. Some readers may notice an odd resemblance between this and the earlier case, in that the super-atoms have a finite number of discrete orientations, just as the atoms do. This resemblance is, however, so superficial and (probably) misleading, that I might not even mention it if I could be sure that it had not been

observed. To state two points of difference among many: the "permitted" directions for the super-atoms depend upon the crystal structure, those for the atoms depend upon the field-direction and the angular momentum of the atom; and if one applies a field to a single crystal in any direction oblique to all of the cubic axes, the super-atoms will consent to point in that direction, provided the field strength is rather high.

Now I must explain what these super-atoms are, since our understanding of them is one of the most satisfactory features in our, as a whole very imperfect, theory of ferromagnetism. They are groups—commonly called *domains*—of adjacent individual atoms; the member-atoms of each domain are behaving like the atoms of a paramagnetic solid. A diagram of a ferromagnetic solid might be drawn as an assemblage of large arrows, each representing the magnetic moment of a single domain; then, around and beside each of these large arrows might be drawn a lot of small arrows representing the magnetic moments of the individual atoms constituting the group; the big arrow would be the resultant of all the little ones. It would not be practicable to do this accurately, for there would have to be millions, or millions of millions, of little arrows to each of the big ones; but even a few suffice to show the idea. It may, however, be recalled that I have lately said that the atoms of a paramagnetic body have an irrestible urge to be in a state of random orientation whenever there is no applied field acting upon them. The resultant of all the little arrows of a domain should then be zero. How can it have a magnitude which is not merely different from zero, but (on the scale customary for such things) very considerable, and independent of the field strength which is applied to the iron?

The answer to this question is given, and very well given, by that extra field or "Weiss field" within the group, which I first mentioned in connection with the constant θ which paramagnetic solids exhibit. It will be remembered how this constant is explained by assuming that the torque, which acts on any one of the atomic magnets, is due not entirely to the applied field H but to the resultant of that and an extra field AI which is proportional to the magnetization I of the body. We have already had the equation (10) which links I and H when this extra field is present. Now striking H altogether out of that equation, we arrive at this one:

$$I = N\mu \tanh(\mu AI/kT) \qquad (12)$$

which refers to a situation in which there is no applied field at all. This may be regarded as an equation for I, fixing the value or values

of I which can exist in this situation. Now everything which I have said so far encourages the reader to suppose that the only possible value of I in the situation is zero; and as a matter of fact, zero is always a solution of this equation. But suppose that there should be another solution, different from zero. The equation would then assert, that if somehow that value of magnetization should arise in the substance, then the extra field would also arise, and in just the right magnitude to maintain that magnetization perpetually, without any aid in the form of a field applied from the outside.

Well, the equation is not exactly easy to solve for I, but it can be mastered—most conveniently by a graphical way—and the striking result is reached, that if T is greater than θ there is no other solution than $I = 0$, but if T is less than θ there *is* a second solution. I will denote this other by I_w. Consider, then, the situation when there is no applied field: if the temperature is higher than θ, I repeat what I have been saying all along, that random orientation of the atomic magnets is inevitable; but when the temperature is lower than θ, then there is another possibility: there is a stable alignment of the atomic magnets entailing this value I_w of the magnetization, which can maintain itself indefinitely if it should ever come into being. Do not leap to the other extreme, and suppose that this is a *perfect* alignment of the atomic magnets and hence a perfect saturation of the domain. Such a situation could exist (according to the theory) only at absolute zero. The equation gives us I_w as function of T, and this function declines smoothly from the value $N\mu$ (for a domain of unit volume!) at absolute zero, to the value zero at $T = \theta$. The curve between these two points is completely determined by the values of μ and θ, which are derived in such ways as I have indicated from the magnetic properties of the substance at the higher temperatures well above θ.

And now, the culmination. The so-called saturation of iron—the ordinate of the I-vs-H curve when it flattens out and becomes sensibly parallel to the axis of abscissae—is itself (as I mentioned) a function of temperature; *it is this same function* (Fig. 3). What is usually called "saturation" with ferromagnetic bodies consists in aligning the big arrows of the domains, so that in unison of direction they exhibit that value of magnetization which is dictated by their internal temperature and internal field. "True" saturation—"saturation of saturations" —the alignment of the atoms within each domain superposed on the alignment of the domains with the field—this can be attained only at the absolute zero of temperature. We are able, however, to work at temperatures so close to absolute zero, that the remaining degree of extrapolation is slight; and we are able, therefore, to give with much

confidence values for the true saturation of iron, nickel, cobalt, gadolinium, and many ferromagnetic alloys.

(The reader may properly wonder why, instead of solving equation (12) obtained by putting $H = 0$ in equation (10), it is not the practice to put for H the field strengths actually applied to iron when aligning

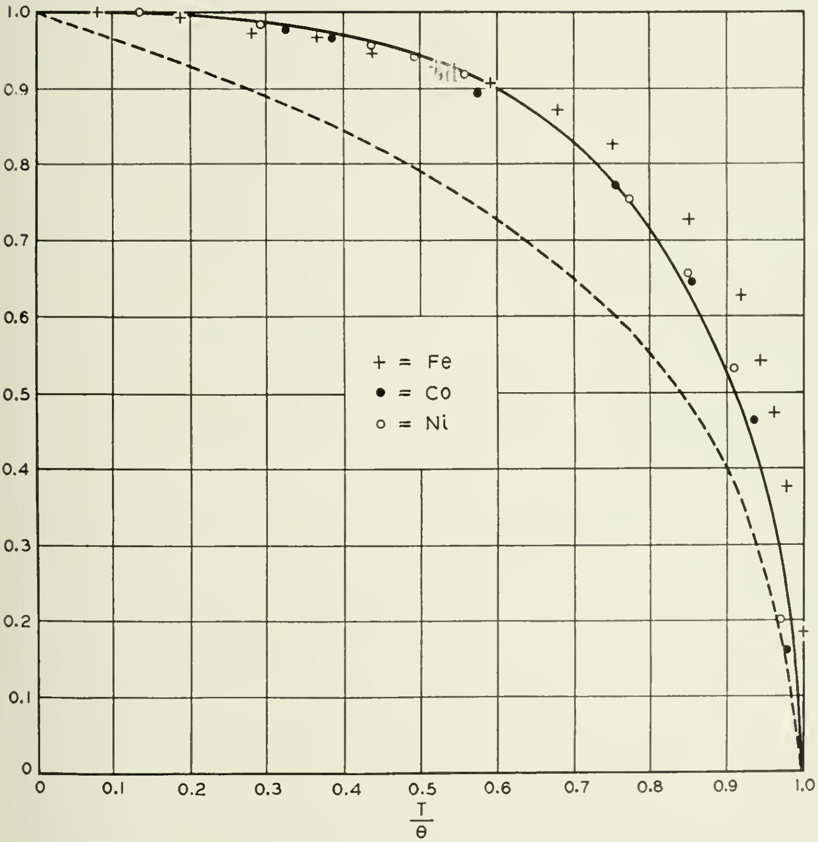


Fig. 3—Intrinsic magnetization plotted against T/θ, for the domains of three ferromagnetic elementary metals (the constant θ has different values for the three). The ordinate is I_w referred to its saturation-value (deduced by extrapolation) as unity. The curves are theoretical, the dashed one by classical theory ($n = \infty$), the full one by quantum-theory ($n = 2$).

the domains or in any other circumstances, and to solve the equation (10) under these conditions? This of course is the correct procedure, but in ferromagnetic bodies AI is usually so enormous by comparison with H, that the latter may be disregarded without appreciable error.)

One naturally asks about the size and the magnetic moment of the domains. It is useless to remember how the latter was determined for paramagnetic bodies from the features of their *I*-vs-*H* curves, since the theory which made that possible is not applicable here. Moreover, the super-atoms share with ordinary atoms the quality of being invisible: no feature of the ordinary surface of a metal indicates them, and no technique of etching the surface seems able to delineate them. (It must be said, however, that ferromagnetic powders sprinkled over ferromagnetic metals may distribute themselves in remarkable picturesque patterns, and perhaps these sometimes simulate the pattern of the underlying domains.*) But fortunately the super-atoms are not inaudible; at least, it is not a very extravagant statement to say that they can be heard. Let a girdle of wire around a rod of some ferromagnetic substance be connected through an amplifier with a microphone, and let a gradually-increasing magnetic field act lengthwise on the rod: the microphone will then emit a machine-gun patter of sharp clicks (with suitable amplification it may be very dramatic!) each of which corresponds to the sudden shift of the magnetic moment or "big arrow" of a domain from one of its possible orientations to another. Now if an electrical instead of an acoustical device is attached to the girdle of wire, the magnitude of the moment which thus re-orients itself at a single click may be assessed. It turns out that the moments are of very various magnitudes; a mean may, however, be estimated, and this mean is some 10^{15} times as great as the moment of a single atom. Therefore the average domain comprises a million billions of atoms, and must therefore be about .002 cm in breadth; but there is a wide range of sizes about the average. As for the individual atoms of the ferromagnetic metals, their moments may be derived from equating $N\mu$ to the values (obtained by extrapolation from observations at various low temperatures, to absolute zero) of that "saturation of saturations" defined above. They are by no means out of the common. Iron and its congeners are readily magnetizable, not because their atoms are extraordinarily magnetic—which is not at all the case—but because their atoms have this curious propensity of cohering together in large groups, developed to an extraordinary degree.

To many features of ferromagnetism, of which whole monographs might be or have been written, I can give only brief mention or none at all. There are the "magneto-caloric effects," arising because, when a ferromagnetic body is heated, the dis-alignment of the atoms

* Cf. the article of R. M. Bozorth in the preceding number (January 1936) of this Journal.

within each domain increases, and this increase requires additional heat over and above that which goes to augment the kinetic energy of the atoms. The specific heat of iron (as of its congeners) is greater than it would be, but for this effect; the excess may be computed from the foregoing theory as function of temperature, and the computed values agree with the data to an extent which speaks very strongly for the theory. (The like is the case with a paramagnetic body exposed to a magnetic field; and as a result, such a body will grow cooler when the applied field is withdrawn, the kinetic energy of the atoms being levied upon when the dis-alignment occurs. The effect is imperceptible in usual circumstances, but with such substances as iron-ammonium alum at liquid-helium temperatures, it becomes so strong that the lowest temperatures ever achieved have been attained by making use of it.) There are the "magnetostrictive effects," arising because, when the atoms of the domains change their orientation, the metal as a whole is strained. It follows that there are interrelations between magnetization, strain, and stress; and anyone remembering even a little of the mathematical theory of elasticity with its moduli and its stress-strain tensors will readily believe that the theory of these interrelations is marvelously complicated. As one sensational example of the consequences, I cite the fact that when a certain permalloy is exposed to a field of, say, one half of one gauss, its magnetization ranges between a few per cent and nearly one hundred per cent of saturation, according to the strength of the tensile stress applied to it. The many processes of the metallurgical arts have often vast effects upon the magnetic properties of the ferromagnetic metals exposed to them: some are due to the changes in the elasticity and hence in the magnetostrictive effects, some to the changes in the chemical constitution (e.g. in the proportion of impurities), some to the changes in phase (of alloys) which these processes entail; but it would be risky to affirm that they have all been traced to one or another of these causes. The finer details in the shape of the I-vs-H curve for ferromagnetics remain to be explained, and to account for one of them it seems to be thought necessary to assume that the domains may gain or lose in size at one another's expense; it is too bad that this impairs the concept of the domain as an immutable super-atom. I leave without overmuch regret this infinitely detailed and complicated topic, to conclude by brief allusions to the spinning electron and to diamagnetism.

Hitherto in these pages I have let it be inferred that when we obtain the magnetic moment of the atom of some element or the molecules of some compound by magnetic experiments upon the substance, it

always agrees with the "theoretical" value derived from the spectrum of that substance when a gas. This is indeed the case with gases and even with a certain number of solids, a large enough number to inspire confidence in the theory. There are, however, numerous exceptions among solids—a circumstance not to be wondered at, since an atom incorporated in a solid is usually in a very different condition from an atom freely wandering about in a gas. The like is true about that number n, the "number of permitted orientations of the atom in a field," which was introduced near the beginning of the article. Either the trend of the I-vs-H curve for a paramagnetic, or the trend of the I_w-vs-H curve for a ferromagnetic, enables us (if it has been sufficiently well measured) to ascertain the value of n; and in a surprising number of instances, comprising iron, cobalt and nickel as well as various rare-earth elements in chemical compounds, the curves prescribe the value *two*, when the free atom according to its spectrum would display some other value. Thus when the atoms are compacted together into a solid, their proximity affects them in such a way as to bring about this result.

Now the important point about this value *two* for n is, that it is the value to be expected for an electron which is either isolated, or else linked to its atom in such a way that it has no orbital angular momentum. The contemporary theory of spectra includes, as one of its essential elements, the postulate of the "spinning electron"—the postulate that each electron by itself is endowed with an intrinsic and indestructible angular momentum and magnetic moment, of definite known amounts, having nothing whatever to do with its orbital revolutions.[4] This angular momentum or "electron-spin" is of the amount which requires $n = 2$, when it is not compounded with an angular momentum of orbital motion or with angular momenta of other electrons. The atoms in question behave, when compacted into solids, as though this angular momentum of individual spinning electrons were the only one left outstanding.

This striking inference is greatly strengthened by measurements upon the one phenomenon in which that angular momentum, which according to atomic theory is always the companion of magnetic moment, comes to light. Imagine a cylinder of some paramagnetic or ferromagnetic substance, hanging freely from a suspension attached to one end. Suppose it to be unmagnetized at first; this signifies that the atoms (whether or not they are grouped into domains) are so oriented that the resultant of all their angular momenta, as well as

[4] The reasons furnished by spectroscopy for making this postulate are much too complex to be interpolated in this article: I refer to the first fourteen pages of "Contemporary Advances in Physics," XXIX, this *Journal*, 14, 285–321 (April 1935).

that of all their magnetic moments, are zero. Now suppose that a field is suddenly applied, parallel to the axis of the cylinder. The substance is suddenly magnetized; this signifies that the resultant of the magnetic moments, and hence that of the angular momenta, are no longer zero. Let I stand (as heretofore) for the former resultant, and P for the latter. Now it is desirable to remember that each atom consists of a nucleus[5] and an electron-family; that the

ERRATA: CONTEMPORARY ADVANCES IN PHYSICS. XXX—THE THEORY OF MAGNETISM—Karl K. Darrow

Bell System Technical Journal, April, 1936

Page 245: Last sentence, " Its lowest possible value (from theory) is $e/2mc$, in which e, m, and c have their usual meanings;"

should read

" One pre-eminent value (from theory) is $e/2mc$, in which e, m, and c have their usual meanings;"

Page 246: First sentence, " Its highest possible value is twice as great;"

should read

" Another pre-eminent value is twice as great;"

This ratio I/P — Its reciprocal is called the "gyromagnetic ratio" is a rare sort of thing: it is a quantity of which the numerical value, measured on pieces of bulk matter, is appropriate also to the elementary particles. If the substance is made up of identical elementary magnets of magnetic moment μ and angular momentum p, then I/P is μ/p. Since μ and p are knowable from spectra, so also is their ratio. Its lowest possible value (from theory) is $e/2mc$, in which e, m, and c have their usual meanings;[6] this would always occur if the electrons had no spins; actually it occurs if the electron-family of the atom is so

[5] Most nuclei possess magnetic moments, which, however, are so excessively small that they can be detected only by experiments of extreme delicacy.

[6] Charge (in E.S.U.) and mass of the electron, and speed of light in vacuo. For the theory underlying these statements, c.f. *l.c.* pp. 285–300. Often the ratio of the experimental value of μ/p to the quantity $e/2mc$ is called an "experimental g-value," the ratio of the theoretical value to $e/2mc$ being conventionally denoted by g.

always agrees with the "theoretical" value derived from the spectrum of that substance when a gas. This is indeed the case with gases and even with a certain number of solids, a large enough number to inspire confidence in the theory. There are, however, numerous exceptions among solids—a circumstance not to be wondered at, since an atom incorporated in a solid is usually in a very different condition from an atom freely wandering about in a gas. The like is true about that num-

other electrons. The atoms in question behave, when compacted into solids, as though this angular momentum of individual spinning electrons were the only one left outstanding.

This striking inference is greatly strengthened by measurements upon the one phenomenon in which that angular momentum, which according to atomic theory is always the companion of magnetic moment, comes to light. Imagine a cylinder of some paramagnetic or ferromagnetic substance, hanging freely from a suspension attached to one end. Suppose it to be unmagnetized at first; this signifies that the atoms (whether or not they are grouped into domains) are so oriented that the resultant of all their angular momenta, as well as

[4] The reasons furnished by spectroscopy for making this postulate are much too complex to be interpolated in this article: I refer to the first fourteen pages of "Contemporary Advances in Physics," XXIX, this *Journal*, 14, 285–321 (April 1935).

that of all their magnetic moments, are zero. Now suppose that a field is suddenly applied, parallel to the axis of the cylinder. The substance is suddenly magnetized; this signifies that the resultant of the magnetic moments, and hence that of the angular momenta, are no longer zero. Let I stand (as heretofore) for the former resultant, and P for the latter. Now it is desirable to remember that each atom consists of a nucleus and an electron-family; that the electron-family possesses the magnetic moment and is oriented in the field (it strictly is what I have hitherto referred to as "the atom"); that the nuclei of the atoms in the cylinder are relatively non-magnetic but contain nearly all of the mass of the cylinder.[5] At the moment of magnetization, the ensemble of the electron-families acquires a net angular momentum P. Now angular momentum being one of these things (like energy and linear momentum) of which the total in Nature does not vary, an equal and opposite amount, $-P$, must appear somewhere or other. It appears in the mass of the cylinder, presumably because of some interaction between the electron-families and their nuclei. The cylinder makes a sharp turn at the instant of magnetization, twisting the suspension from which it hangs through an angle from which (and from the rigidity of the suspension) the value of $-P$ can be found. This effect and its converse (an unmagnetized cylinder may be magnetized by sharply twisting it) are known as the "gyromagnetic effects." They are delicate and difficult to produce, *a fortiori* to measure; yet of late years experimenters have succeeded in measuring P together with I, and therefore learning the value of the ratio I/P—first for the ferromagnetic metals and then for some of their compounds and alloys, and lately for certain paramagnetic salts, the work on these last being done at the very low temperatures where alone they can be strongly magnetized.

This ratio I/P—its reciprocal is called the "gyromagnetic ratio"—is a rare sort of thing: it is a quantity of which the numerical value, measured on pieces of bulk matter, is appropriate also to the elementary particles. If the substance is made up of identical elementary magnets of magnetic moment μ and angular momentum p, then I/P is μ/p. Since μ and p are knowable from spectra, so also is their ratio. Its lowest possible value (from theory) is $e/2mc$, in which e, m, and c have their usual meanings;[6] this would always occur if the electrons had no spins; actually it occurs if the electron-family of the atom is so

[5] Most nuclei possess magnetic moments, which, however, are so excessively small that they can be detected only by experiments of extreme delicacy.

[6] Charge (in E.S.U.) and mass of the electron, and speed of light in vacuo. For the theory underlying these statements, c.f. *l.c.* pp. 285–300. Often the ratio of the experimental value of μ/p to the quantity $e/2mc$ is called an "experimental g-value," the ratio of the theoretical value to $e/2mc$ being conventionally denoted by g.

organized that the spins neutralize one another. Its highest possible value is twice as great; this occurs if nothing counts excepting the electron-spins, and signifies either that the electrons are free [7] or else that the electron-family of each atom is so organized that there is no net angular momentum due to orbital motion. Intermediate values are possible and signify different types of organization of the electron-family. The values predicted from spectra have been confirmed for a few of the rare-earth atoms in their paramagnetic salts; but usually, as I have already intimated, the *observed value of the ratio* μ/p is about $2(e/2mc)$, though the spectrum says something else.

It would be pleasant now to add that the magnetic moment of each of these substances, per atom, amounts to some integer multiple of the magnetic moment μ_e of the spinning electron. We then could say that the integer is the number of "uncompensated" spinning electrons in the atom, implying by the word "uncompensated" in this connection that all the magnetic moments in the electron-family of the atom add up vectorially to zero and so do all the angular momenta, with the sole exception of those pertaining to these electron-spins. Such is not, however, the case: some of the experimental values are $2.2\mu_e$ for iron, $1.7\mu_e$ for cobalt, $0.6\mu_e$ for nickel. It seems necessary to assume that in metallic solid iron, some of the atoms present two uncompensated electrons to the orienting field, and others three. Iron in different chemical compounds exhibits different values of magnetic moment, and sometimes the ratio μ/p is different from $2(e/2mc)$, suggesting that angular momenta of orbital motion are not quite cancelled out; indeed it now appears that the ratio is slightly but definitely different from this specific value even in the cases (such as those of the pure ferromagnetic metals and of permalloy) in which at first the measurements suggested that it was the same.

Such observations as these last are problems for the specialists in atomic theory; magnetism offers great numbers of these problems. Another and a complementary way of viewing this situation is, to look on every measurement of a magnetic moment made upon a solid as an item of information about an atom (or a molecule) existing in a condition which is not accessible to spectroscopic research. Spectra indicate the normal state of atoms in freedom; occasional magnetic experiments (like those on gaseous oxygen here cited, or those on molecular beams by the Gerlach-Stern method, which I hope to treat on a later occasion) also refer to free atoms and molecules, and confirm the indications of the spectra, thus sustaining both the methods; but

[7] Certain metals, the alkali metals for instance, exhibit a paramagnetism which is entirely due to the "free" or conduction electrons.

mostly the magnetic methods refer to atoms in a solid, and so they make available a new and broad domain for the operations of atomic theory.

There remains diamagnetism. The first thing to be said about the theory of diamagnetism is discouraging; for it has the earmark of a futile atomic theory—it involves the assumption that the individual atoms behave exactly like the substance as a whole. Under all field strengths and all conditions, it is assumed that the diamagnetic moment of a block of N atoms is N times the diamagnetic moment of a single atom. Nevertheless this is not a futile assumption, for strictly it is not an assumption at all but an inference from atomic structure. It was mentioned early in these pages that owing to the unbreakable link between angular momentum and magnetic moment, a magnetic atom precesses about the direction of the field. This motion of precession is an extra motion of the electrons of the atoms, a circulatory motion around the axis supplied by the direction of the field. This extra motion entails an extra current, which entails an extra magnetic moment, which is the source of diamagnetism or which *is* diamagnetism. *Diamagnetism is precession.* It is not confined, as the foregoing words suggest, to atoms which have a net magnetic moment. Consider an atom (a free atom of any noble gas will afford an example) possessing two or more electrons, the orbits and the spins of which are so oriented that the resultant magnetic moment is nil. Though some of the orbits and spins are pointed oppositely to others, they all precess in the same sense, and the atom acquires a magnetic moment in the field though it had none beforehand. The like is true, of course, when the resultant of the orbits and the spins is different from zero; the agents of orientation which were discussed above render it paramagnetic, but the precession renders it diamagnetic, and it is paramagnetic and diamagnetic—or ferromagnetic and diamagnetic—at one and the same time. The moment due to the precession is proportional to the field strength, and the factor of proportionality may be calculated from the structure of the atom (it depends primarily upon the areas of the electron-orbits). The agreement of the calculated values with the data is generally satisfactory; and diamagnetism, the least conspicuous of the three types of magnetism, takes precedence over the others as being that one of the three of which our understanding is most nearly perfect.

The Proportioning of Shielded Circuits for Minimum High-Frequency Attenuation

By E. I. GREEN, F. A. LEIBE and H. E. CURTIS

For given conditions of design there exists an optimum proportioning or configuration which makes the high-frequency attenuation of a given type of individually shielded circuit a minimum. Determination is made of such optimum proportioning for a wide variety of types of individually shielded circuits including several novel types designed to make the high-frequency attenuation low in comparison with the cross-sectional area occupied by the circuit, and the attenuation of different types is compared. The following topics and specific circuit structures are considered:

COAXIAL CIRCUITS—Basic Coaxial Circuit; Effect of Dielectric; Effect of Frequency on Optimum Ratio; Thin Walls; Stranded Conductors; Optimum Proportioning as a Function of Conductor Resistance.

BALANCED SHIELDED CIRCUITS—Shielded Pair (Cylindrical Conductors and Shield)—Condition for Minimum Attenuation, Condition for Maximum Characteristic Impedance, Effect of Dielectric, Effect of Frequency; Pair in Space; Shielded Stranded Pair; Pair with Shield Return; Double Coaxial Circuit; Shielded Pair (Round Conductors and Oval Shield); Shielded Pair (Quasi-Elliptical Conductors); Shielded Quad.

INTRODUCTION

SINCE the very beginning of mathematics, problems of maximizing and minimizing have possessed a marked fascination. The Greeks were successful in solving a few geometric problems of this character. Later, algebra was found to be another method of attack. Finally, the powerful methods of the calculus became available for the determination of maxima and minima in manifold variety. The reasons for the continued interest in such problems are not hard to find. It is but natural to seek the ideal, and here, at least, is one phase of mankind's search for perfection in which a goodly measure of success may be achieved. In addition, a knowledge of the optimum dimensioning of things, or of the optimum relations between things, frequently holds much practical significance.

It is mainly with problems of maxima and minima that this paper is concerned. These problems have to do with transmission circuits which are surrounded by individual shields. Recent literature [1, 2] has pointed out that circuits of this type have properties which render them especially suitable for the transmission of broad bands of frequencies. Such circuits are also finding application as "lead-ins" to connect radio antennas with transmitting or receiving apparatus.[3, 4]

[1] For numbered references, see end of paper.

It is well at this juncture to understand the function of shielding in a high-frequency transmission circuit. Such shielding serves one or both of these purposes: (*a*) keeping interference due to external sources from entering the circuit, and (*b*) preventing the circuit from causing interference in external circuits. The shielding may either supplement or completely replace the use of electrical balance to reduce interference. The design of shield, that is, its construction, material, thickness, etc., is determined by the degree of shielding required and by considerations of mechanical performance and cost. The degree of shielding needed depends in turn upon such factors as the type and length of circuit, the nature and frequency of the signals to be transmitted, and the magnitudes of external interference. These interesting aspects of shield design, some of which have been dealt with elsewhere,[1, 2, 5] will not be discussed here.

Attention will rather be directed to an intriguing property of any individually shielded circuit, namely, that, for given conditions of design, there always exists an optimum proportioning or configuration which makes the transmission efficiency of the circuit a maximum, or, in other words, makes the attenuation a minimum. One such condition of design which may be imposed is that the cross-sectional area enclosed within the shield is to be a constant. In what follows, determination will be made of such optimum proportioning for a wide variety of types of individually shielded circuits. Since the attenuation is generally of outstanding importance in a high-frequency transmission line, the results should be not only of theoretical interest but also of practical value. Moreover, the different methods which are used in solving these problems should find further application, both in the many other known problems which must perforce be omitted for lack of space, and in those problems which may be conceived in the future.

The principal types of individually shielded circuits to be discussed are:

(1) Coaxial or concentric circuits, in which an outer conductor, which serves also as a shield, completely surrounds a centrally disposed inner conductor.

(2) Shielded pairs, consisting of a pair of conductors which form the transmission circuit, these being surrounded by an individual conducting shield.

The coaxial circuit is unbalanced, and relies solely upon shielding for protection against interference from or into its exterior. In contrast to this is the balanced type of circuit, in which the go and return

conductors are designed to be substantially alike and are located substantially symmetrically with respect to earth and surrounding conductors.

In the past, telephone transmission circuits have been largely of the balanced type. It has been found possible to operate such balanced circuits up to fairly high frequencies,[2] without incurring excessive interference. However, as the frequency is raised it becomes increasingly difficult to maintain a sufficiently high degree of balance, and shielding may then be desirable. The shielding may eliminate balance entirely, as in the coaxial circuit, or may be combined with balance in what may be termed a shielded balanced circuit, of which the shielded pair is an outstanding example.

For the simplest forms of circuits, the optimum relations may be precisely derived with the aid of the propagation formulas. In more difficult cases it is necessary to use approximate methods of one kind or another. These methods, however, can generally be made to yield sufficiently accurate results for practical purposes.

Coaxial Circuits

Coaxial circuits, which furnish the least difficult problems in optimum proportioning, make a natural starting point for this subject.

Basic Coaxial Circuit

The first type of circuit to be considered is the basic circuit consisting of two tubular conductors arranged coaxially, whose cross-section is shown diagrammatically in Fig. 1.

Before trying to find out how to proportion such a circuit, it must be noted that in the design of any shielded circuit there enter a number of variables, including the overall size of the structure, the type and thickness of shield, the type of conductor or conductors, the type of insulation, and the frequencies to be transmitted. Some of these factors exert an important influence on the optimum proportioning, so that it is necessary, in order to arrive at a unique solution in a given case, to keep certain factors fixed. Thereafter, however, the effect produced upon the result by varying these factors may be examined.

First, therefore, let the following assumptions be made:

1. That the tubular conductors of Fig. 1 are composed of solid material.
2. That the dielectric is gaseous, with zero dielectric loss. This is a condition which may be approached in practice.
3. That the inner diameter of the outer conductor is fixed. This is a convenient assumption, having for its basis the fact that it is ordinarily desirable, for economic or other reasons, to limit the

size of the outer conductor, and the further fact that the thickness of the outer conductor will ordinarily be determined by mechanical considerations or by shielding requirements.

4. That the frequency is high enough to permit the use of certain approximate formulas as noted below. Practically, this means that at the frequency considered the currents are largely crowded toward the inner surface of the outer conductor and the outer surface of the inner conductor.

The problem is to discover the proportioning which will make the high-frequency attenuation of the circuit a minimum under such conditions. It is well known that the attenuation of a transmission

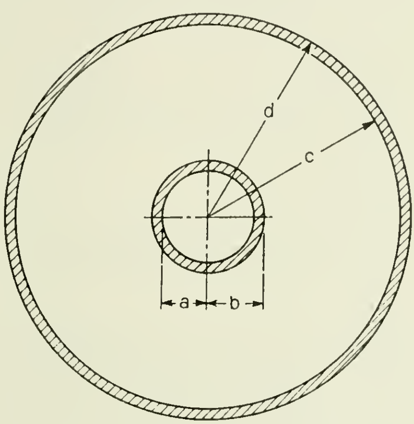


Fig. 1—Coaxial conductor circuit.

circuit at high frequencies may be represented by the following approximate formula:[6]

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \text{ nepers per cm.,} \tag{1}$$

where R, L, G and C designate, respectively, the linear resistance, inductance, conductance and capacitance of the circuit. Except as otherwise indicated, values in this and subsequent formulas are expressed in c.g.s. electromagnetic units.

When the dielectric loss is negligible, the second term of formula (1) evidently disappears.

Let a and b represent, respectively, the inner and outer radii of the inner conductor, c and d the inner and outer radii of the outer conductor, f the frequency, λ_1 and μ_1, respectively, the conductivity and

permeability of the material of the inner conductor, and λ_2 and μ_2 the corresponding values for the outer conductor. The ratio λ_1/λ_2 will be designated by n.

The high-frequency resistance of the inner conductor may then be approximately expressed by the formula:[5, 7]

$$R_1 = \frac{1}{b}\sqrt{\frac{f\mu_1}{\lambda_1}} \text{ abohms per cm.} \tag{2}$$

Similarly the high-frequency resistance of the outer conductor is approximately:

$$R_0 = \frac{1}{c}\sqrt{\frac{f\mu_2}{\lambda_2}} \text{ abohms per cm.} \tag{3}$$

The high-frequency inductance of the circuit is approximately [7]

$$L = 2 \log_e \frac{c}{b} \text{ abhenries per cm.} \tag{4}$$

The capacitance of the circuit is [8]

$$C = \frac{\epsilon}{2 \log_e \frac{c}{b}} \text{ abfarads per cm.,} \tag{5}$$

where ϵ is the dielectric constant of the dielectric material between conductors, equal to $1/9 \times 10^{-20}$ for gaseous dielectric, corresponding to unity in the practical system of units.

The high-frequency attenuation of the coaxial circuit with negligible dielectric loss, obtained by combining the above formulas, is

$$\alpha = \frac{1}{2c}\sqrt{\frac{f}{\lambda_1}}\left(\frac{c}{b} + \sqrt{n}\right)\frac{\sqrt{\epsilon}}{2 \log_e \frac{c}{b}} \text{ nepers per cm.} \tag{6}$$

The value of permeability assumed in the above equation, and hereafter, is unity, but the methods may be used also for other values.

If the inner diameter of outer conductor be assumed fixed, this expression may be minimized with respect to the ratio c/b, which is the ratio of the radii (or diameters). For convenience this ratio may be designated as ρ. It is found that the high-frequency attenuation is a minimum when the value of ρ is that given by

$$\log_e \rho = \frac{\rho + \sqrt{n}}{\rho}. \tag{7}$$

Figure 2 shows the values of the ratio ρ which satisfy this relation plotted as a function of the conductivity ratio n.

It is noteworthy that the optimum ratio of radii or diameters is independent of (*a*) the diameter and thickness of outer conductor, (*b*) the inner diameter of the inner conductor, and (*c*) the frequency, provided the frequency is high enough for the approximate formulas to hold. It follows from (*a*) that, assuming a fixed thickness of outer conductor, moderately small in comparison with its diameter, relation (7) makes it possible to find the minimum size of outer conductor with which a given value of high-frequency attenuation may be realized. It follows from (*b*) that the inner conductor may be either hollow or solid, provided that the approximate resistance formulas are valid.

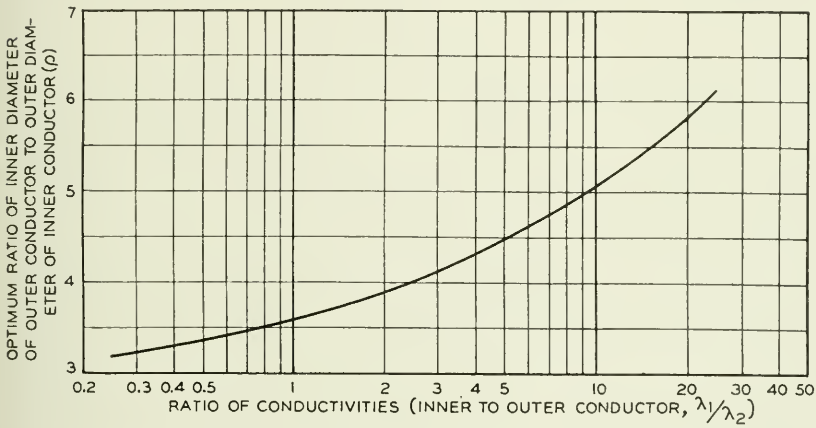


Fig. 2—Variation of optimum diameter ratio of coaxial circuit with conductivity ratio.

A case of special interest arises when the two conductors have the same conductivity, that is, when n equals unity. For this condition the solution of (7) is *

$$\rho = \frac{c}{b} = 3.59. \tag{8}$$

A practical example of the case of different conductivities is a coaxial structure in which the inner conductor is of copper and the outer conductor of lead. For a lead outer conductor containing about 1 per cent of antimony, the ratio of conductivities of inner and outer

* The existence of an optimum relation of this kind was first noted by C. S. Franklin, who gave the value as 3.7. (See Reference 3.) Subsequently the precise value was derived independently of Franklin. (See Reference 10.)

conductors is approximately 13, and the optimum diameter ratio for such a structure, as found from Fig. 2, is about 5.25.

The behavior of the attenuation in the vicinity of the optimum diameter ratio is illustrated in Fig. 3, which shows attenuation plotted

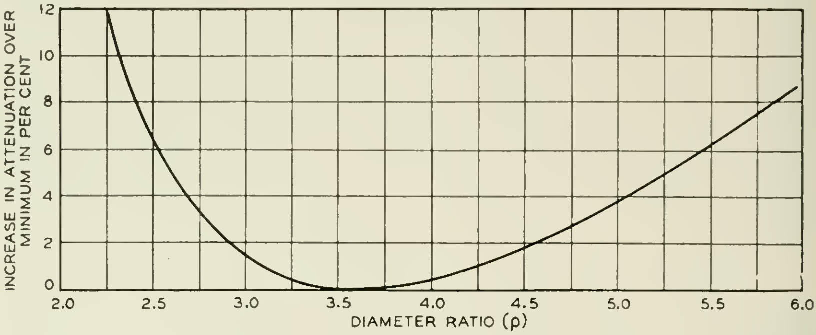


Fig. 3—Variation of optimum diameter ratio of coaxial circuit with conductivity ratio.

against diameter ratio for the case where n equals unity. It will be seen that near the optimum the attenuation changes very slowly. This is fortunate, since it means that unavoidable departures from the optimum diameter ratio may be permitted without appreciable effect on the attenuation. Other small departures from ideal design are also allowable. Thus, for example, it has been assumed in deriving the condition for minimum attenuation that the two conductors of the circuit are perfectly coaxial or concentric. However, for moderately small departures from perfect concentricity occasioned by practical difficulties of construction, the conditions for minimum attenuation are substantially the same as for a circuit with no eccentricity. The situation is similar for other types of shielded circuits to be considered later, in these cases also only a reasonably close approximation to the ideal being necessary.

Effect of Dielectric

Suppose now that the capacitance and leakage conductance introduced by the insulation are substantial.[9] First, it will be assumed that the space between the two conductors is filled with a substantially uniform non-gaseous dielectric material having a dielectric constant ϵ and a power factor p. Such would be the case, for example, if the two coaxial conductors were separated by a continuous rubber insulation. The leakage conductance of the circuit now becomes

$$G = p\omega C = \frac{p\omega\epsilon}{2 \log_\epsilon \frac{c}{b}} \text{ abmhos per cm.,} \tag{9}$$

where, as usual, ω equals $2\pi f$.

By substituting in formula (1), the high-frequency attenuation is found to be

$$\alpha = \frac{1}{2c}\sqrt{\frac{f\epsilon}{\lambda_1}}(\rho + \sqrt{n})\frac{1}{2\log_e \rho} + \frac{p\omega\sqrt{\epsilon}}{2} \text{ nepers per cm.} \tag{10}$$

Since ω, p, and ϵ are not functions of the ratio c/b, the second term of this expression is constant for purposes of differentiation with respect to that ratio, and the condition for minimum attenuation is identical with that previously found, as given in formula (7).

A high-frequency transmission property of smaller interest than the attenuation is the characteristic impedance. This is given by the familiar formula [6]

$$Z_0 = \sqrt{\frac{L}{C}} \text{ abohms.} \tag{11}$$

For the coaxial circuit with dielectric constant ϵ the high-frequency characteristic impedance is

$$Z_0 = \frac{2\log_e \rho}{\sqrt{\epsilon}} \text{ abohms.} \tag{12}$$

There now comes the case where the space between the conductors consists of a combination of gaseous and non-gaseous dielectrics. Perhaps the simplest example occurs when the conductors are separated by insulating discs or washers extending continuously between the two conductors with flat sides perpendicular thereto. Such a construction is illustrated in Fig. 4. Let the thickness of each insulating

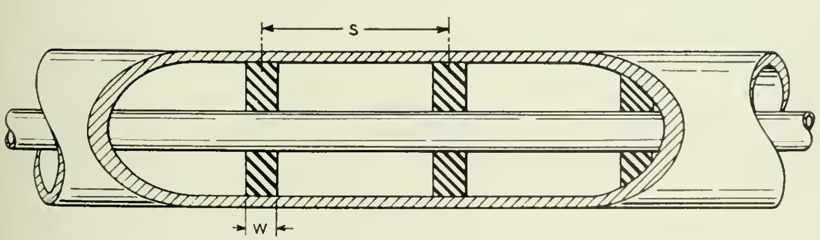


Fig. 4—Coaxial structure with disc insulation.

disc be designated w, the spacing between centers of adjacent discs, s, the dielectric constant of the air dielectric, ϵ_1, and that of the disc material ϵ_2.

The capacitance of the coaxial circuit now becomes

$$C = \frac{\epsilon_1 + \dfrac{(\epsilon_2 - \epsilon_1)w}{s}}{2 \log_e \rho} \text{ abfarads per cm.,} \tag{13}$$

while the leakage conductance is

$$G = \frac{w}{s} \cdot \frac{p\omega\epsilon_2}{2 \log_e \rho} \text{ abmhos per cm.} \tag{14}$$

On substituting these values in formula (1) the following expression results:

$$\alpha = \frac{1}{2c}\sqrt{\frac{f}{\lambda_1}} \frac{\sqrt{\epsilon_1 s + (\epsilon_2 - \epsilon_1)w}}{\sqrt{s}} \frac{\rho + \sqrt{n}}{2 \log_e \rho} + \frac{p\omega\epsilon_2}{2} \frac{w}{\sqrt{s}} \frac{1}{\sqrt{\epsilon_1 s + (\epsilon_2 - \epsilon_1)w}} \text{ nepers per cm.} \tag{15}$$

Once more the second term is independent of c and b, and the condition for minimum attenuation is, as before, that given by equation (7).

The high-frequency characteristic impedance in this case, however, is

$$Z_0 = \frac{2 \log_e \rho}{\sqrt{\epsilon_1 + \dfrac{(\epsilon_2 - \epsilon_1)w}{s}}} \text{ abohms.} \tag{16}$$

The quantity in the denominator of the above expression is evidently the weighted average dielectric constant of the insulating medium.

In the case just considered, the gaseous and non-gaseous dielectrics were separated from each other by planes perpendicular to the axis of the conductors. Consequently, each line of dielectric flux passed through only one kind of material. It can be shown that, as long as this latter condition holds, the condition for minimum high-frequency attenuation as given by equation (7) is valid, or, in other words, the optimum diameter ratio is that shown in Fig. 2. Cases arise, however, in which a line of dielectric flux, in going from one conductor to the other, may pass through more than one kind of dielectric material. It is extremely difficult to obtain a mathematical solution for the diameter ratio which results in minimum attenuation for such cases, since this involves a three-dimensional field problem. Consideration of the problem, however, indicates that the optimum diameter ratio will not differ appreciably from that given by Fig. 2, especially if the dielectric is mostly gaseous, which, of course, is highly desirable.

Effect of Frequency on Optimum Ratio

It has been seen that, at the higher frequencies where the approximate transmission formulas may be employed, the optimum diameter ratio is substantially independent of frequency. In so far as the practical application of individually shielded circuits is concerned, it is in these higher frequencies that interest primarily centers. Even when it is desired to transmit a wide band extending from high frequencies down to comparatively low ones, it is advantageous to proportion the circuit so as to minimize the attenuation at the highest transmitted frequency, since the attenuation at all lower frequencies will be less than the value thus obtained.

It may, however, be worth while to consider briefly the question of optimum proportioning when low frequencies only are involved. The appropriate transmission formulas to be used instead of the approximate high-frequency expressions are known,[5] and the optimum diameter ratio in any specific case may be derived from these. It will be evident that, since skin effect is present to a lesser degree at the low frequencies, the diameter and thickness of the outer conductor and the thickness of the inner conductor will, as the frequency is decreased, have an increasing influence on the optimum proportioning.

Without attempting to derive precise values for the different conditions, it may be noted that the optimum diameter ratio for low frequencies is invariably less than that for high frequencies, the high-frequency value being approached asymptotically as a limit. The reason for this will be readily apparent. Let the inner diameter and thickness of the outer conductor be assumed fixed. At high frequencies the resistance of the inner conductor varies inversely with the first power of its diameter. At lower frequencies, however, this resistance varies inversely with some power of the diameter greater than unity, and finally, at zero frequency, assuming a solid wire, with the square of the diameter. Hence it is, that, in varying the size of the inner conductor in order to obtain a balance between the change of resistance and change of capacity, it is advantageous to make the inner conductor somewhat larger, or, in other words, to make the diameter ratio smaller, at low frequencies than at high frequencies.

Thin Walls

What is the result if the walls of the two coaxial conductors are made very thin? Under this condition the conductor resistance, and hence the attenuation, will remain substantially constant over a wide range of frequencies. This constancy is realized, however, at the expense of an increase in the attenuation as compared with that for thicker conductor walls.

Using the notation of Fig. 1, the resistances of the inner and outer conductors, both with conductivity λ, at frequencies where the walls are sufficiently thin to avoid skin effect, are

$$R_i = \frac{1}{\pi\lambda(b^2 - a^2)} \text{ abohms per cm.} \tag{17}$$

$$R_0 = \frac{1}{\pi\lambda(d^2 - c^2)} \text{ abohms per cm.} \tag{18}$$

Let the inner conductor have a fixed thickness $b - a$, the outer conductor a thickness $d - c$, and let the ratio $(b - a)/(d - c)$ be represented by t. For small values of wall thickness

$$b^2 - a^2 \doteqdot 2b(b - a) \tag{19}$$

and

$$d^2 - c^2 \doteqdot 2c(d - c) = 2c\frac{(b - a)}{t}. \tag{20}$$

Substituting these relations and the values of L and C from (4) and (5) in equation (1), it is found that the attenuation for the circuit with thin walls is

$$\alpha = \frac{\sqrt{\epsilon}}{4\pi\lambda c}\frac{(\rho + t)}{(b - a)}\frac{1}{2\log_e \rho} \text{ nepers per cm.} \tag{21}$$

Differentiation shows that minimum attenuation in the case of thin walls is obtained when

$$\log_e \rho = \frac{\rho + t}{\rho}. \tag{22}$$

The values of diameter ratio which satisfy this relation may be found from the curve of Fig. 2, if the values of abscissæ on that curve are interpreted as values of t^2.

If the conductor walls are thin, as above, and if in addition the conductivities of the two conductors are not the same, that of the inner conductor being n times that of the outer one, the condition for minimum attenuation becomes

$$\log_e \rho = \frac{\rho + nt}{\rho}. \tag{23}$$

Figure 2 may be used to find the values of diameter ratio which satisfy this relation also, the abscissæ scale markings in this case being taken as values of n^2t^2.

Stranded Conductors

With conductors having solid walls, or composed of non-insulated strips or filaments, the currents at high frequencies are largely crowded toward the inner surface of the outer conductor and the outer surface of the inner conductor, due to skin effect. Since the losses in the conductors themselves ordinarily comprise the major portion of the attenuation in a coaxial circuit, interest attaches to the possibility of counteracting the increase in conductor resistance due to skin effect by using a conductor composed of a number of individually insulated strands so twisted or interwoven as to distribute the current more nearly uniformly over the cross-section.[11] Chief attention naturally focuses upon the inner conductor, which is by far the greater contributor to the resistance, and this discussion will be largely limited to the case where only the inner coaxial conductor is stranded.*

Types of stranded conductors suitable for use as the inner conductor of a coaxial circuit include both those in which the conductor cross-section is completely filled with insulated strands and those in which the insulated strands form an annular cross-section, surrounding a core of non-conducting or conducting material. Of various possible methods of stranding, one simple and effective process is similar to that used in the construction of rope. A few strands are twisted together to form a group, several such groups are twisted into a larger group, and so on until the desired conductor cross-section is obtained.

The high-frequency resistance of a stranded conductor may be determined either by measurement or computation. For a completely stranded inner conductor of any diameter, size, number of strands, and thickness of insulation, the high-frequency resistance is given by S. Butterworth [12] and in unpublished material by J. R. Carson. The resistance values obtained in measurements of stranded conductors approximate very closely the theoretical results.

In evaluating the results obtained with stranding, it is convenient to compare the resistance of a stranded conductor with that of a non-stranded conductor of the same overall size. For the case of a stranded inner conductor, the ratio of the resistance of the stranded conductor at any given frequency to the resistance at the same frequency of a solid conductor having the same outer diameter and composed of the same material used in the strands may be designated as m.

The values of the resistance ratio m which may be realized in practice depend upon the frequency and the design of stranded conductor. Some idea of these values for two specific conductors may be obtained

* "Stranded" is used to mean "composed of insulated strands."

from the curves of Fig. 5. It will be seen that there is ordinarily a frequency at which the resistance ratio is a minimum. Above this frequency the improvement due to stranding rapidly vanishes, the performance thereafter being worse than that of the corresponding non-stranded conductor. The minimum value of resistance ratio attained in the range of some hundreds of kilocycles may be in the order of 0.6, a very substantial improvement. In order to secure any marked advantage in the frequency range above 700 or 800 kilocycles, the number and fineness of the individual strands would be such as practically to preclude their use.

Another result obtained with stranding is an increase in the internal inductance of the conductors, which likewise serves to reduce the high-

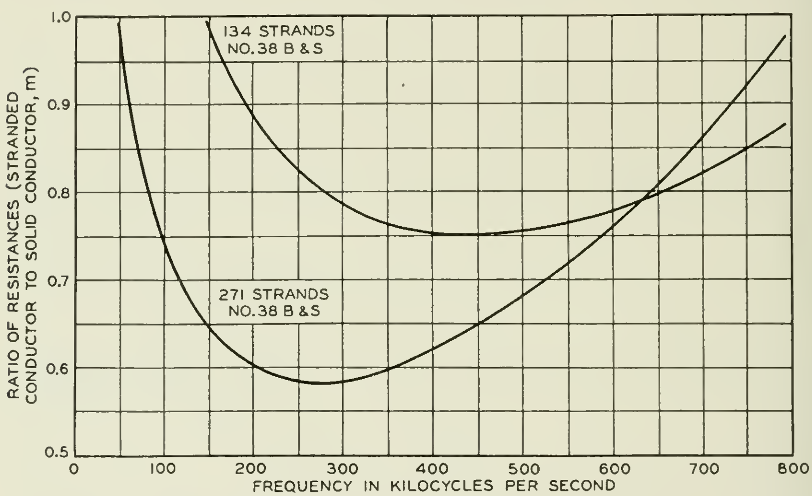


Fig. 5—Resistance ratios of stranded conductors.

frequency attenuation. For a round conductor which is completely stranded, the internal inductance at all frequencies where the current is uniformly distributed over the conductor cross-section approximates .5 abhenry per centimeter, which is the internal inductance of a solid round wire at zero frequency. In general, this value of internal inductance will hold up to frequencies somewhat above that for which the resistance ratio m is a minimum. The internal inductance of a stranded conductor of annular cross-section, for all frequencies where the current is uniformly distributed over the cross-section, is

$$L_i = \frac{b^2 - 3a^2}{2(b^2 - a^2)} + \frac{2a^4}{(b^2 - a^2)^2} \log_e \frac{b}{a} \text{ abhenries per cm.} \qquad (24)$$

This is the same as the internal inductance at zero frequency of a solid tube of the same dimensions.

Since either the inner or outer conductor of a coaxial circuit, or both, may be stranded, and since, in addition, the dielectric loss may either be negligible or may be appreciable, there are six different cases of optimum proportioning which might be considered.[13] Only one case, however, that of a coaxial circuit with only the inner conductor stranded and with negligible dielectric loss, will be taken up here. The high-frequency attenuation of such a coaxial circuit is

$$\alpha = \frac{m}{2c}\sqrt{\frac{f}{\lambda_1}}\left(\rho + \frac{\sqrt{n}}{m}\right)\sqrt{\frac{\epsilon}{(4\log_e \rho)^2 + 2L_i \log_e \rho}} \text{ nepers per cm.} \qquad (25)$$

While the value of m varies with frequency and with the design of the stranded conductor, this value is, for a particular frequency and a particular design, definitely determinable. As has been noted, it is generally desirable to proportion a transmission circuit so as to minimize the attenuation at the highest frequency to be transmitted. Furthermore, the value of m will not vary rapidly with changes in conductor diameter provided the number of strands be changed as the conductor size is varied. It therefore becomes possible to treat m as a constant in deriving the relation for optimum proportioning.

Using ρ to designate c/b, the condition for minimum high-frequency attenuation is found to be

$$\frac{2\dfrac{m}{\sqrt{n}}\rho\log_e\rho}{\dfrac{m}{\sqrt{n}}\rho + 1} = \frac{4\log_e\rho + L_1}{2\log_e\rho + L_1}. \qquad (26)$$

Figure 6 shows graphs of equation (26) for two values of L_i, namely, $L_i = 0.5$ abhenry per centimeter, which corresponds to the case where the cross-section of the inner conductor is completely stranded, and $L_i = 0$. When the stranded inner conductor is of annular cross-section the optimum value of the diameter ratio lies somewhere between the two curves shown. The useful range of m probably lies between about 0.5 and unity and that of n between about 1 and 15.

As to the practical use of stranding, it is apparent from the resistance ratio curves of Fig. 5 that in order to take advantage of stranding it would be necessary to limit the transmission band to a maximum frequency well below that possible with non-stranded conductors. Further drawbacks to the use of stranded conductors are their greater cost as compared with non-stranded ones, and greater mechanical

difficulties in using them. For these reasons stranded conductors do not seem likely to find early application in broad band transmission circuits.

Optimum Proportioning as a Function of Conductor Resistance

The optimum diameter ratio of a coaxial circuit may also be expressed broadly as a function of the two conductor resistances. Assume a coaxial circuit in which the high-frequency resistance of the inner conductor varies, at least over a limited range, inversely as its

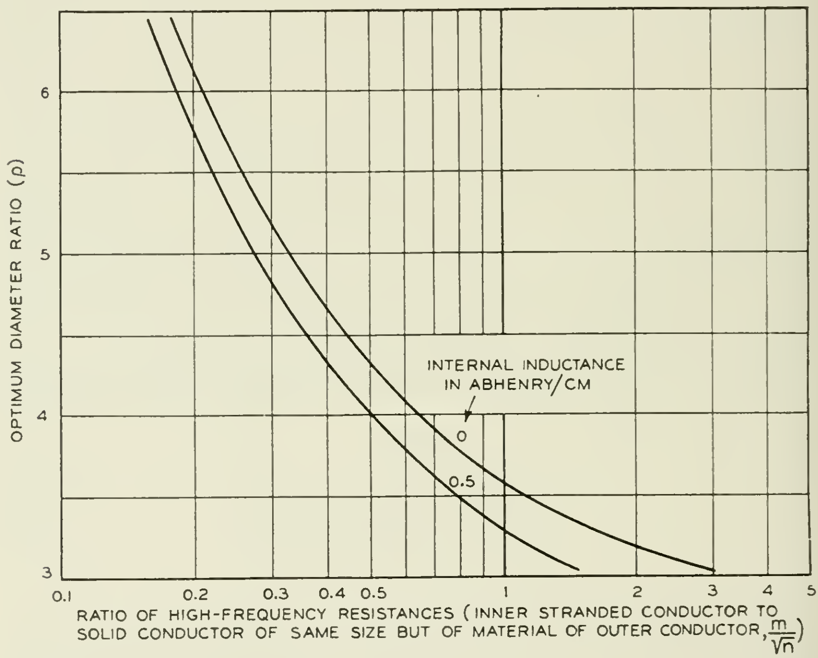


Fig. 6—Optimum diameter ratio of coaxial circuit with stranded inner conductor.

outer radius and that of the outer conductor as its inner radius, thus

$$R_i = \frac{k_1}{b} \qquad \text{and} \qquad R_0 = \frac{k_2}{c}. \tag{27), (28}$$

These relations are approximately true for all the types of circuits which have been discussed. Let

$$r = \frac{k_2}{k_1} = \frac{R_0}{R_i} \cdot \frac{c}{b} = \frac{R_0}{R_i}\rho. \tag{29}$$

If the internal inductance of the conductors is assumed to be zero, which is the most usual case, the high-frequency attenuation of the circuit may then be written

$$\alpha = \frac{k_1}{2c}(\rho + r)\frac{1}{2 \log_e \rho}. \tag{30}$$

Upon minimizing with respect to c/b the condition for minimum high-frequency attenuation is found to be

$$\log_e \rho = \frac{\rho + r}{\rho} = 1 + \frac{R_0}{R_i}. \tag{31}$$

These relations have been found useful in certain instances.

Balanced Shielded Circuits

Though arrangements of three or more coaxial conductors are possible,[14] practical interest is almost wholly limited to coaxial circuits employing but two conductors. With balanced shielded circuits, however, the number of conductors, counting the shield as one, is necessarily three and may be more. With a coaxial circuit, moreover, the cylindrical shape is the natural and usual one for the conductors. With balanced shielded circuits, on the other hand, there enter a number of possibilities. Not only are cylindrical shapes of conductors and shield to be considered, but a variety of other shapes as well. More complex, therefore, than the foregoing problems in optimum proportioning are those for balanced shielded circuits, now to be discussed.

Shielded Pair—Cylindrical Conductors and Shield

The simplest form of balanced shielded circuit is a shielded pair comprising two cylindrical conductors surrounded by a cylindrical shield. Such a circuit is shown diagrammatically in cross-section in Fig. 7. For the present, attention will be directed to the circuit obtained when the two enclosed conductors are connected one as a return for the other.

Condition for Minimum Attenuation [15]

As before, it is desired to minimize the high-frequency attenuation. Let it be assumed first, as in the coaxial circuit, that the area within the shield is fixed, the conductors are of solid material and the dielectric is gaseous. Let b represent the radius of each conductor in Fig. 7, c the inner radius of the shield, h the distance from the center of either conductor to the center of the shield, λ_1 the conductivity of each con-

ductor, λ_2 that of the shield, and n the ratio of λ_1/λ_2. Expressions for the high-frequency attenuation of this circuit have been given in unpublished formulas developed by S. A. Schelkunoff and by Mrs. S. P. Mead. The approximate formula given below is due to the latter.

$$\alpha = \frac{\rho\left[1 + \dfrac{1+2\nu^2}{4\nu^4}(1 - 4\sigma^2)\right] + 4\sqrt{n}\,\sigma^2\left[1 + \sigma^4 - \dfrac{1+4\nu^2}{8\nu^4}\right]}{\log_e\left[2\nu\dfrac{1-\sigma^2}{1+\sigma^2}\right] - \dfrac{1+4\nu^2}{16\nu^4}(1 - 4\sigma^2)} \times \frac{1}{4c}\frac{\sqrt{f\epsilon}}{\sqrt{\lambda_1}} \text{ nepers per cm.,} \qquad (32)$$

where $\sigma = h/c$ and $\nu = h/b$.

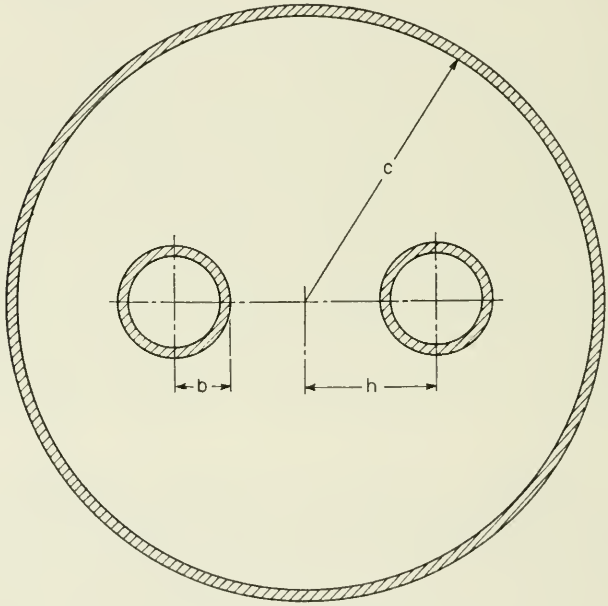


Fig. 7—Shielded pair.

The values of the diameter ratio (ρ) and what may be termed the spacing ratio (σ), which make this expression a minimum for different values of the conductivity ratio n, can be determined in different ways. One possible method is to find the values of h and b which satisfy the equations $\partial\alpha/\partial h = 0$ and $\partial\alpha/\partial b = 0$. The partial derivatives are, however, very complicated. Accordingly a preferable alternative is to substitute various pairs of values of ρ and σ in (32) and determine, graphically or otherwise, the particular pair which makes it a minimum. In this way it is found that when the conductors and shield are of the

same material, so that n equals unity, the optimum values are approximately

$$\rho = \frac{c}{b} = 5.4; \qquad \sigma = \frac{h}{c} = .46. \qquad (33), (34)$$

The optimum diameter and spacing ratios for different values of the conductivity ratio n are shown in Figs. 8 and 9. For copper conduc-

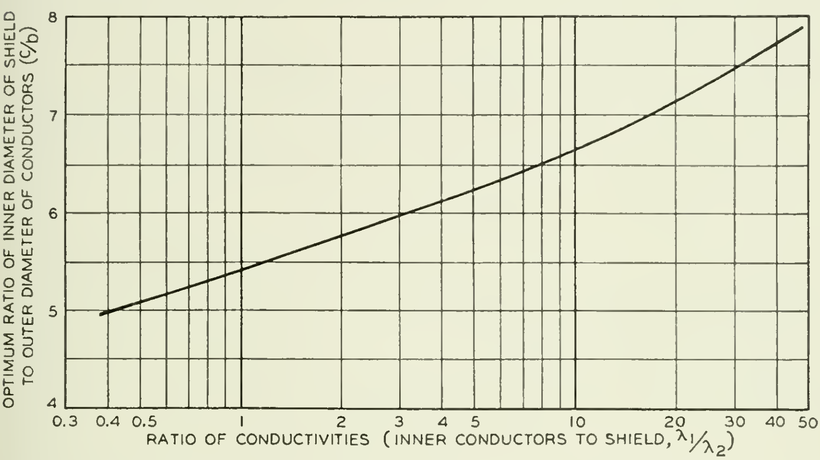


Fig. 8—Variation of optimum diameter ratio of shielded pair with conductivity ratio.

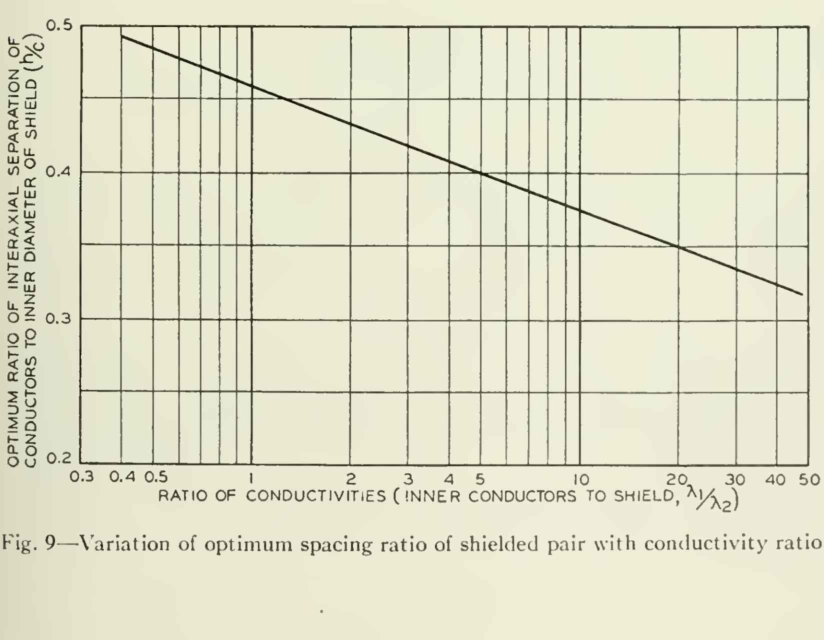


Fig. 9—Variation of optimum spacing ratio of shielded pair with conductivity ratio.

tors and a lead shield, the values are approximately 6.9 and .36, respectively.

As with the coaxial circuit, these optimum relations are independent of the diameter and thickness of the shield. Hence they make it possible to find the minimum size of shield necessary for a given value of high-frequency attenuation. The optimum relations are also independent of the frequency, provided the frequency is high enough for the approximate formulas to hold. The inner conductors may be either hollow or solid.

Condition for Maximum Characteristic Impedance[15]

Occasionally it is of interest to know the condition that must be satisfied to obtain maximum high-frequency characteristic impedance for a solid pair with circular shield. At high frequencies the value of $1/\sqrt{LC}$ approaches a constant value equal to the velocity of light divided by the square root of the ratio of the dielectric constant of the circuit to that of air. Hence the condition for maximum characteristic impedance is also, from equation (11), that for maximum inductance and minimum capacitance.

Accordingly, the high-frequency characteristic impedance of the shielded solid pair circuit is given by the formula:

$$Z_0 = \frac{4}{\sqrt{\epsilon}}\left(\log_e\left[2\nu \frac{1 - \sigma^2}{1 + \sigma^2} \right] - \frac{1 + 4\nu^2}{16\nu^4}(1 - 4\sigma^2) \right) \text{abohms.} \quad (35)$$

Let it be assumed first that the wires are very small compared with the shield. Then equation (35) may be written

$$Z_0 = \frac{4}{\sqrt{\epsilon}} \log_e \left[\sigma \frac{1 - \sigma^2}{1 + \sigma^2} \right] + \frac{4}{\sqrt{\epsilon}} \log_e 2p \text{ abohms.} \quad (36)$$

For a given ratio of inner diameter of shield to outer diameter of conductor, the second term of this expression is constant. By minimizing the first term with respect to σ, it is found that, so long as the ratio of inner diameter of shield to conductor diameter is large, maximum characteristic impedance is obtained when σ has a value of .486.

If the conductors are large compared with the shield, equation (36) no longer holds. However, since the capacitance and high-frequency characteristic impedance are inversely proportional to one another, the position of the conductors with respect to the shield must be such as to minimize the capacitance. It is clear that as the conductor diameter approaches the inner radius of the shield, σ approaches 0.5 for minimum capacitance. Hence, for any ratio of inner diameter of shield to

diameter of conductor, the ratio of the interaxial separation of the conductors to the inner diameter of the shield which gives maximum characteristic impedance lies between the limits 0.486 and 0.500. For practical purposes a value of about 0.49 may generally be used.

Effect of Dielectric

The effect of dielectric for a shielded pair is similar to that for a coaxial circuit. When the insulation is so disposed between conductors and shield that a line of dielectric flux passes through only one kind of dielectric material, the second term of the attenuation formula is independent of the proportioning of conductors and shield, so that the optimum proportions as given in Figs. 8 and 9 are unchanged. These values will also serve for most practical cases where a line of dielectric flux may pass through more than one kind of material.

Effect of Frequency

At frequencies where the approximate formulas no longer hold, the conditions for minimum attenuation as given by Figs. 8 and 9 undergo some change, especially the former. As the frequency is decreased the attenuation is minimized by increasing the size of conductor for a given size of shield. In other words, the optimum diameter ratio grows less. The optimum spacing ratio increases from 0.46 toward the value which gives minimum capacitance, i.e., approximately 0.49.

Pair in Space

It is interesting to digress for a moment to consider briefly the case shown in Fig. 10 of a pair of round conductors in space. This may

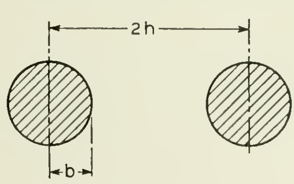


Fig. 10—Pair in space.

be regarded as a pair surrounded by a shield of infinite diameter. If the conductors are of solid material, the attenuation of the circuit at high frequencies is

$$\alpha = \frac{P}{4b}\sqrt{\frac{f\epsilon}{\lambda}}\frac{1}{\cosh^{-1}\nu} \text{ nepers per cm.,} \tag{37}$$

where P is the proximity effect factor, given in a paper by J. R. Carson.[16] At high frequencies this factor reduces to the asymptotic value

$$P = \frac{\nu}{\sqrt{\nu^2 - 1}} \cdot \tag{38}$$

For a given high frequency and given wire separation, assuming the dielectric constant and conductivity to be fixed, equation (37) becomes

$$\alpha = \frac{K_3 \nu^2}{\sqrt{\nu^2 - 1} \cosh^{-1} \nu}, \tag{39}$$

where K_3 is a constant.

For a given wire separation this expression is minimized when

$$\nu = \frac{h}{b} = 2.27. \tag{49}$$

For open-wire pairs, which may be considered as approaching pairs in space, it is ordinarily cheaper to obtain any desired attenuation at a given frequency by using a wide separation and relatively small conductors rather than a narrow separation and conductors of such size as to satisfy (40). This relation is of considerable utility, however, in that it is a reasonably close approximation to the optimum for many kinds of shielded pairs. The corresponding ratio for the shielded solid pair, as given by (33) and (34), is approximately 2.5.

Shielded Stranded Pair

The preceding discussion of shielded pairs has been limited to types of enclosed conductors such that high-frequency currents are crowded toward the conductor surfaces. There will now be found the optimum proportioning when the enclosed conductors are stranded.[17]

The capacitance and inductance between two shielded stranded wires when surrounded by a cylindrical shield are approximately

$$C = \frac{\epsilon}{4 \log_e \left[2\nu \dfrac{1 - \sigma^2}{1 + \sigma^2} \right]} \text{ abfarads per cm.,} \tag{41}$$

$$L = 4 \log_e \left[2\nu \frac{1 - \sigma^2}{1 + \sigma^2} \right] + 2L_i \text{ abhenries per cm.,} \tag{42}$$

where L_i is the internal inductance of each conductor.

If it be assumed that the current distribution is uniform over the cross-section of the enclosed conductors, the resistance of each is the

same as if its return were coaxial. Hence the high-frequency resistance of each conductor is

$$R_i = \frac{2m}{b}\sqrt{\frac{f}{\lambda_1}} \text{ abohms per cm.} \tag{43}$$

The high-frequency resistance of the shield can be shown to be

$$R_0 = \frac{8ch^2}{c^4 - h^4}\sqrt{\frac{f}{\lambda_2}} \text{ abohms per cm.} \tag{44}$$

The high-frequency attenuation of the shielded stranded pair, found by substituting equations (41) to (44) in (1), is, with zero dielectric loss,

$$\alpha = \frac{\dfrac{m}{2c}\sqrt{\dfrac{f}{\lambda_1}}\left[\rho + \dfrac{4\sqrt{n}\sigma^2}{m(1 - \sigma^4)}\right]}{\sqrt{\left[\log_e 2\nu\dfrac{1 - \sigma^2}{1 + \sigma^2}\right]\left[4\log_e 2\nu\dfrac{1 - \sigma^2}{1 + \sigma^2} + 2L_i\right]}} \text{ nepers per cm.} \tag{45}$$

The optimum proportions of the shielded stranded pair at high frequencies depend, therefore, on the two quantities $m/\sqrt{n}$ and L_i. For any given shield radius c, the values of h and b which give minimum attenuation may be found by setting

$$\frac{\partial\alpha}{\partial h} = 0; \qquad \text{and} \qquad \frac{\partial\alpha}{\partial b} = 0. \tag{46), (47}$$

By imposing the first condition it is found that

$$\frac{\partial M}{\partial h}\frac{(c^4 - h^4)^2}{8ch(c^4h^4)} = M\frac{2\log_e M(4\log_e M + 2L_i)}{8\log_e M + 2L_i}\cdot\frac{1}{\dfrac{m}{\sqrt{n}b} + \dfrac{4ch^2}{c^4 - h^4}}\cdot \tag{48}$$

Imposing the second condition we find that

$$-\frac{\sqrt{n}b^2}{m}\frac{M}{b} = M\frac{2\log_e M(4\log_e M + 2L_i)}{8\log_e M + 2L_i}\cdot\frac{1}{\dfrac{m}{\sqrt{n}b} + \dfrac{4ch^2}{c^4 - h^4}}, \tag{49}$$

where $M = 2\nu(1 - \sigma^2)/(1 + \sigma^2)$.

Upon equating the left hand members of (48) and (49), and substituting the values of the derivatives, the following expression results

$$\rho = \frac{8\sigma^2(1 + \sigma^4)}{\dfrac{m}{\sqrt{n}}(1 - \sigma^4)(1 - 4\sigma^2 - \sigma^4)}\cdot \tag{50}$$

This expression is the locus of values of the ratio ρ which give minimum attenuation for different assumed values of the ratio σ. The unique values of $h/c = \sigma$ and $c/b = \rho$, which give minimum attenuation for a given value of $m/\sqrt{n}$ and L_i, may be obtained by taking pairs of σ and ρ which satisfy equation (50), substituting them in equation (45), and graphically determining the pair for which the attenuation is a minimum.

Figures 11 and 12 show a graph, obtained in this way, of the optimum

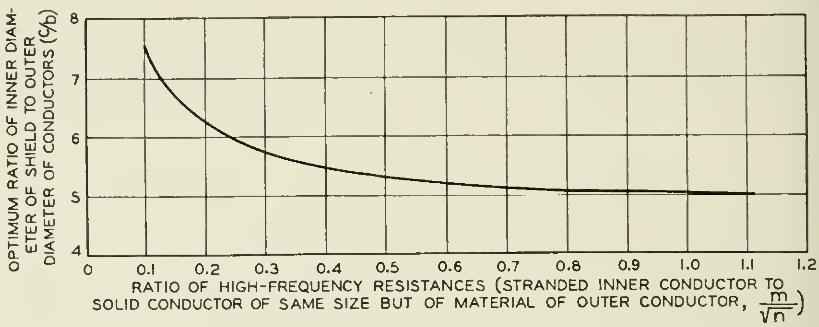


Fig. 11—Optimum diameter ratio of shielded stranded pair.

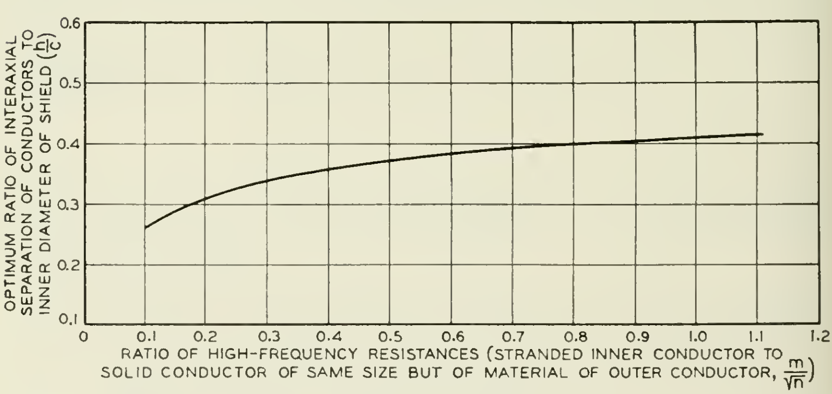


Fig. 12—Optimum spacing ratio of shielded stranded pair.

proportions for a shielded stranded pair, plotted as a function of $m/\sqrt{n}$ for a value of L_i equal to 0.5 abhenry per centimeter, which corresponds to the case where each conductor is completely stranded.

Pair with Shield Return

The discussion of the shielded pair thus far has been concerned solely with the circuit which employs one of the enclosed conductors as a return for the other. A second circuit may be obtained by

transmitting over the two enclosed conductors in parallel with the shield as the return. This latter circuit alone is less efficient than a coaxial circuit formed by replacing the two inner conductors with a single one. If, however, the two circuits obtainable from the shielded pair structure can both be employed without excessive mutual interference, there will be a considerable increase in the usefulness of the system, measured in terms of the total frequency range that can be transmitted without exceeding a given attenuation. It is therefore of interest to determine the conditions making the total transmitted frequency range for the two circuits a maximum.[18]

The high-frequency attenuation of each circuit, assuming solid conductors, can be written

$$\alpha = K\sqrt{f}, \tag{51}$$

where K is a constant, different for each circuit, which depends on the size and material of the conductors, and the dielectric constant of the insulation. Leakage is assumed negligibly small.

Using subscripts 1 and 2, respectively, to designate the circuit comprising the two enclosed conductors one as a return for the other and the circuit comprising the two wires in parallel with shield return, it follows that

$$f_1 + f_2 = \frac{\alpha_1^2}{K_1^2} + \frac{\alpha_2^2}{K_2^2} \cdot \tag{52}$$

Letting $A = \alpha_2/\alpha_1$

$$f_1 + f_2 = \alpha_1^2\left(\frac{1}{K_1^2} + \frac{A^2}{K_2^2}\right) \cdot \tag{53}$$

Equation (53) gives the sum of the frequency ranges that can be transmitted in the above manner over any given shielded pair for any given attenuation at the highest frequencies of the bands. To obtain maximum total range, this equation must be maximized.

The attenuation of the circuit comprising one enclosed conductor as a return for the other is given by equation (32), from which the value of K_1 can be obtained immediately. An expression for K_2 has been given in an unpublished formula due to Mrs. S. P. Mead, as follows:

$$K_2 = \frac{U + V}{\log_e\left[\dfrac{\rho(1 - \sigma^4)}{2\sigma}\right] - \dfrac{1 + 4\sigma^4}{1 + 4\nu^2}\left(1 + 4\sigma^4\left(\dfrac{5 + 4\nu^2}{1 + 4\nu^2}\right)\right)} \frac{1}{4c}\sqrt{\frac{\epsilon}{\lambda_1}} \tag{54}$$

in which

$$U = \rho\left[1 + \frac{8\nu^2(1 + 4\sigma^4)}{(1 + 4\nu^2)^2}\left(1 + 4\sigma^4\left(\frac{9 + 4\nu^2}{1 + 4\nu^2}\right)\right)\right],$$

$$V = 2\left[1 + 2\sigma^4 + \frac{8\sigma^4}{1 + 4\nu^2}\left(1 + 8\sigma^4\left(\frac{5 + 4\nu^2}{1 + 4\nu^2}\right)\right)\right] \cdot$$

For a given inner radius of shield and given dielectric constant and conductor material, the diameter and spacing ratios which make equation (53) a maximum can be obtained by the substitution method previously described. When the conductors and shield are of the same material and $A = 1$, computation shows that the total frequency range is a maximum when the radius ratio ρ equals 5.9 and the spacing ratio σ equals 0.33. This value of $A = 1$ represents an important practical case, since it will, as a rule, be desirable to employ the same repeater points for each circuit and permit the same attenuations between repeater points. It is also of interest, however, to determine the effect of other values of A.

When A is zero, the problem reduces to that of the simple shielded pair, which has been shown previously to be minimized by the proportions given in (33) and (34).

When A becomes large, or, in other words, when the phantom circuit alone is used, $1/K_2^2$ must be maximized. It is obviously necessary that the enclosed conductor be in contact and, accordingly, the spacing ratio must be the reciprocal of the diameter ratio. For this condition the following proportions result:

$$\rho = \frac{c}{b} = 6.0; \qquad \sigma = \frac{h}{c} = 0.17. \tag{55), (56}$$

The above proportions are optimum only when the enclosed conductors and the shield are of the same conductivity. The relations for the case of unequal conductivities may be derived in a similar manner. For practical purposes the effect of dielectric loss on the optimum proportions is negligible.

Double Coaxial Circuit

Another form of balanced and shielded transmission circuits may be obtained by using two coaxial conductor units, the transmission path consisting of the two inner coaxial conductors in series, with the outer coaxial conductors serving only for shielding. Such a circuit is shown diagrammatically in cross-section in Fig. 13. Usually the outer conductors would be in practically continuous contact with each other. A circuit of this type will handle a frequency band extending to lower values than can be used with a single coaxial circuit, since it is balanced and the two coaxial units can be transposed by twisting or by periodic interchange of their positions. At high frequencies, where the shielding of the outer conductor of the coaxial circuit becomes effective, the outer conductors may be separated to any desired distance. It is essential, however, that they be connected together at the ends of the circuit.

In such a double coaxial circuit, used at high frequencies, equal and opposite currents will flow on the inner and outer conductors of each coaxial unit. The resistance and inductance of the balanced circuit will, therefore, be twice, and the capacitance and leakance one-half, the corresponding values for one coaxial unit. The attenuation of the balanced circuit is equal to the attenuation of one coaxial unit and may be expressed by the formulas previously given, where the various symbols are understood to refer to one unit of the circuit. Accordingly the optimum high-frequency proportions are the same as those previously derived for ordinary coaxial circuits of different types.[19]

As the frequency is reduced, the optimum proportions become different from those for coaxial units, since the circuit inductance approximates more closely that for a simple pair of wires occupying the positions of the inner conductors, while the capacitance remains equal to one-half of that of one coaxial unit. As a result the optimum diameter ratio is larger than at high frequencies.

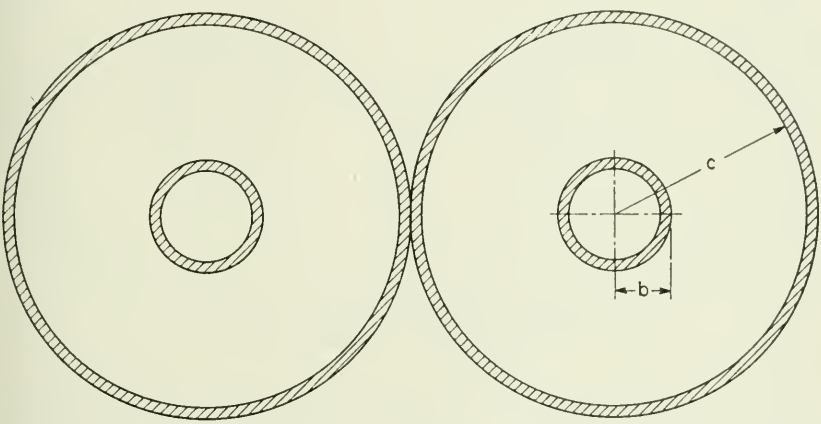


Fig. 13—Double coaxial circuit.

Shielded Pair—Round Conductors and Oval Shield

The shield around a pair does not have to be cylindrical. Upon consideration of a pair of round conductors with a cylindrical shield, as shown in Fig. 7, it is evident that the shield approaches quite close to the conductors at the sides, while it is well removed from them at the top and bottom of the figure. This means that for a given area enclosed by the shield the capacitance of the circuit is greater than would be the case if the shield were kept at a more nearly uniform

distance from the conductors. Consequently, for a given area circumscribed by the shield, a reduction of attenuation can be secured by changing the shape of the shield.

The problem of determining the shape of shield which gives minimum high-frequency attenuation presents extreme difficulty, and a rigorous solution has not been obtained. However, it appears that a close approach to the ideal shape can be obtained by a shield having the cross-section shown in Fig. 14, which consists of two semi-circles

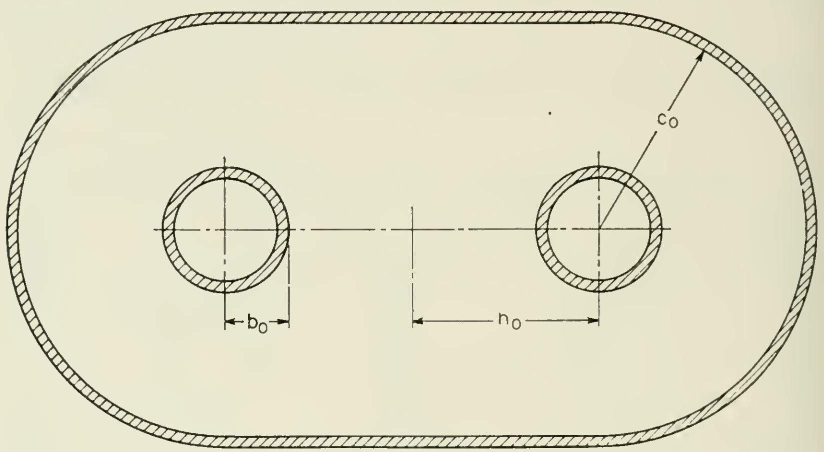


Fig. 14—Oval shielded pair.

joined by straight lines, the inner conductors being placed at the centers of the semi-circles. For convenience this shape of shield will be termed "oval."

The optimum proportioning [20] for a pair of conductors with such an oval shield may be closely approximated by comparison with the pair with circular shield and with the double coaxial circuit. In such comparison the cross-sectional areas of the different circuits will be assumed equal.

Consideration will first be given to the case where the enclosed conductors in Fig. 14 are of solid material. The conductivity of the conductors will be assumed the same as that of the shield, it being apparent that the same methods may be employed in the case of different conductivities. In arriving at the spacing ratio of the conductors for minimum attenuation, the condition for minimum capacitance will be used as a stepping stone. The spacing ratio of the conductors in Fig. 14 may be represented by $h_0/(c_0 + h_0)$. Comparison with Fig. 7 shows that the corresponding ratio for that figure is h/c, which,

it has already been seen, should have a value of approximately .486 for minimum capacitance. It is evident that the value of the ratio $h_0/(c_0 + h_0)$ for minimum capacitance in Fig. 14 should be very close to .486, but in view of the concentricity of the conductors with the semi-circular parts of the shield it should be slightly less than this value. It has been found that to obtain minimum capacitance for an oval shielded circuit the spacing ratio should be approximately

$$\frac{h_0}{c_0 + h_0} \doteq .47. \tag{57}$$

It has been seen for Fig. 7 that to obtain minimum high frequency attenuation the spacing ratio is shifted from the value of .486, which gives minimum capacitance, to a value of about .46. For Fig. 14, however, the current density in the shield is more uniform, so that the proximity effect between conductors is less completely compensated by the shield currents. Hence the spacing ratio for minimum high-frequency attenuation for the oval shielded circuit should be approximately the same as that for minimum capacitance, as given in (57) above.

There remains to be determined the second condition for minimum high-frequency attenuation for an oval shielded circuit of given cross-sectional area, namely, the optimum value of the diameter ratio c_0/b_0. Comparison with Fig. 13 indicates that the optimum value of this ratio should be fairly close to the optimum value of 3.6 for the coaxial circuit. Comparison with Fig. 7, c_0 being equal to about $.69c$ for equal areas in the two cases, shows that the optimum value of the ratio c_0/b_0 should be slightly greater than 3.6. For practical purposes the optimum may be taken as

$$\frac{c_0}{b_0} \doteq 3.7. \tag{58}$$

With this ratio the size of the conductors with oval shield is, for the same cross-sectional area, approximately the same as that of the optimum size of conductors with circular shield.

The capacitance of the pair with oval shield is smaller than the capacitance of the pair with circular shield, because the inner conductors of the former are more widely separated and are farther from the shield. It is very slightly larger than the capacitance of the double coaxial circuit.

The part of the resistance of the oval shielded circuit which is due to the shield will be less than that for a circular shield because of the more uniform current density in the shield. However, as has been

noted, the proximity effect between conductors is less completely neutralized by the shield currents than is the case for the circular shield. It appears that these two effects may approximately balance one another, and that the circuit resistance is approximately the same for both oval and circular shielded circuits.

It is found that a circuit of approximately optimum proportions comprising two solid round wires surrounded by an oval shield has about 12 per cent lower attenuation than a circuit with circular shield of equal cross-sectional area.

When the conductors enclosed within the oval shield are stranded there is no increase of conductor resistance due to proximity effect. On this account it is desirable to bring the conductors closer together in order to reduce the shield loss and the optimum spacing ratio will be less than for the case of solid conductors. With stranded conductors the attenuation reduction as compared with the circular shield is greater than in the case of solid wires; for example, if the resistance ratio (m) is .7, the attenuation with oval shield will be about 25 per cent less than that of the circular shield.

The circular form of shield is ordinarily the most convenient and practical one. A disadvantage of an oval shield as compared thereto is unequal stiffness or resistance to bending in different directions.

Shielded Pair—Quasi-Elliptical Conductors

It has been suggested at different times that the ordinary round form of conductor, while well adapted for manufacturing purposes, may not be the theoretically optimum shape for many types of high-frequency transmission circuits. Speculations in this respect have differed greatly, and a large variety of non-circular shapes of conductors have been proposed, including flat strips, strips with concave or convex faces opposite one another, angular forms, etc. However, except in the case of the coaxial circuit, for which the circular form is clearly the optimum, there has been, so far as the authors are aware, no exact analytical determination of the optimum conductor shape for a given type of circuit.

A complete treatment of possible problems of this kind would extend to great length. It is worth while, however, to consider a single problem, namely, that of determining what shape and spacing for a pair of conductors with circular shield will result in minimum high-frequency attenuation. This problem is of particular interest inasmuch as the circular shape is ordinarily the most convenient and practical one for a shield.

In attacking this problem the fundamental principles which deter-

mine the high-frequency attenuation of a circuit comprising a pair of conductors surrounded by a shield may be briefly examined. At high frequencies, where the currents are crowded toward the surfaces of the conductors, the attenuation is proportional to the product of the resistance and capacitance of the circuit, both of which are functions of the flux density in the dielectric.

With a circular shield, and round conductors, the flux density is far from uniform around the surfaces of the conductors, being relatively high at points nearest the shield and also at points nearest the shield's center, and a minimum at points about half-way between. Accordingly, it appears that the high-frequency resistance of the conductors can be reduced by reshaping them so as to make the flux distribution more uniform. This can be accomplished by squeezing the conductors at regions of maximum flux density and bulging them at regions of minimum flux density, thereby producing a conductor of approximately elliptical cross-section.

The flux distribution around the shield is also far from uniform, being a maximum at points nearest the conductors and a minimum at points 90 degrees away. Making the enclosed conductors elliptical tends to reduce this non-uniformity, thereby reducing the circuit resistance due to loss in the shield.

This process of reshaping the conductors can not be carried very far, however, because it soon increases the circuit capacitance more than it decreases the resistance. It is difficult to treat this problem by rigorous mathematics, but an analysis can be made which yields an approximate solution.

For certain conductor shapes, the high-frequency attenuation of a pair with circular shield may be determined by a method involving the substitution of charged filaments for the conductors. Let any number of positively and negatively charged filaments be included in the shield, the net charge on the filaments being zero. The electrostatic potential at any point of this system can readily be determined by known methods. Thus, for example, Fig. 15 shows the location of the equipotential surfaces for the case of two oppositely charged filaments placed within a circular shield, the distance from each filament to the center of the shield being .46 times the shield radius.

In any such system, a conducting cylinder whose external surface corresponds to, and whose potential is equal to the potential of, a particular equipotential surface may be substituted for the part of the system contained within that surface without disturbing the flux distribution external to it. Consequently, the capacitance of a shielded circuit employing equal and oppositely charged conductors

having the same shape as any two corresponding equipotential surfaces of the electrostatic system can be determined.

The flux density at any points on the conductors or on the shield is proportional to the rate of change of the potential with respect to the normal to the surface at that point. The high-frequency resistances of the conductors and shield, respectively, are proportional to the

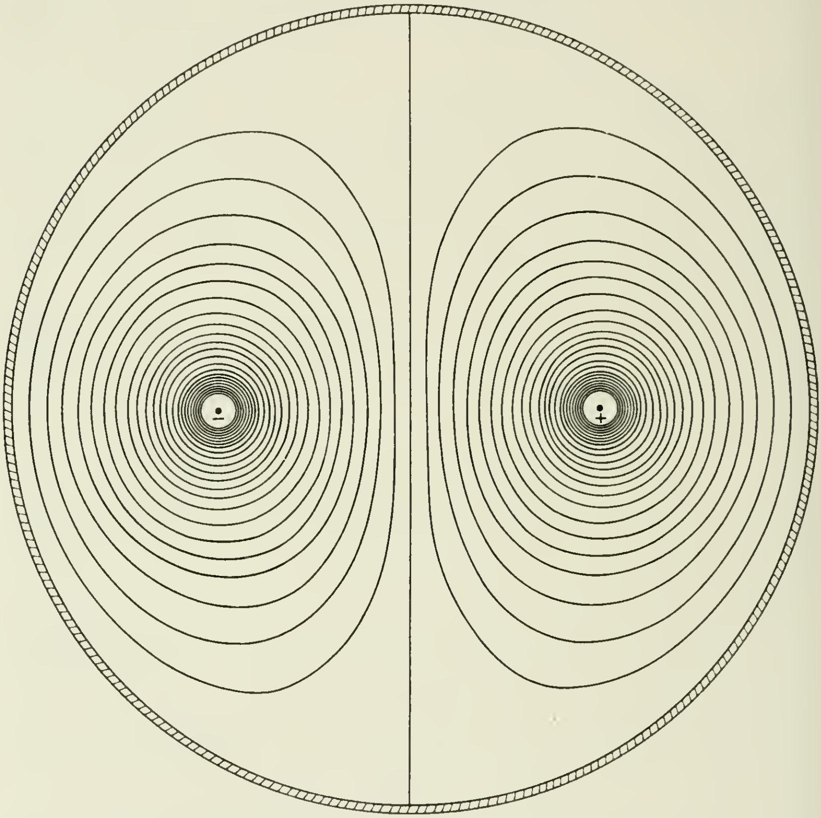


Fig. 15—Equipotential lines around shielded charged filaments.

integral of the square of the flux density around their periphery. Thus the high-frequency resistance of the circuit may be determined, and from this and the capacitance, the high-frequency attenuation.

This method makes it possible to determine and compare the high-frequency attenuations of conductors having shapes corresponding to the equipotential surfaces for various assumed arrangements and numbers of charged filaments. If, however, the problem be that of

determining the attenuation for a given shape of conductor, there may be great difficulty in finding the arrangement and number of filaments which will produce an equipotential surface to coincide with the given shape.

By applying this method to a series of approximately elliptical conductors previously shown to be of the shape that would be expected to have lower attenuation than circular conductors, what is considered a close approximation to the optimum shape of conductor for a pair with circular shield has been arrived at. This is approximately an ellipse whose major axis is about 5 per cent longer than its minor axis, the latter being in line with the center of the shield. The high-frequency attenuation of a circuit with circular shield and conductors of this shape is approximately 2 per cent lower than that for the same shield with round conductors. This reduction does not appear enough to offset the practical difficulties involved with conductors of such shape.

Shielded Quad

The number of conductors enclosed within a shield, instead of being one, as in the coaxial, or two, as in the shielded pair, may be more. By placing four conductors within a common shield, two separate balanced-to-ground circuits may be obtained. If sufficiently good balance can be obtained between these circuits, the total frequency band which can be transmitted within a given cross-sectional area may be increased. To obtain balance, the plane of the conductors of one circuit needs to be at right angles to that of the other circuit and all conductors should be equidistant from the axis of the shield. The pairs may be twisted or spiralled about the axis of the shield.

An arrangement of this kind is shown in Fig. 16, where four round conductors are placed within a circular shield to form a shielded quad, or, as it is frequently described when the conductors are twisted, a "shielded spiral four." Diagonally opposite conductors are used as the sides of a circuit.

Approximate formulas for the high-frequency attenuation of either circuit of Fig. 16, when the enclosed conductors are solid, have been derived in unpublished work of Mrs. S. P. Mead and S. A. Schelkunoff. The optimum high-frequency proportioning of the system, assuming the same conductivity for both enclosed conductors and assuming gaseous dielectric, has been determined by Mrs. Mead. The results are shown in Figs. 17 and 18, where the optimum diameter ratio and spacing ratio are plotted as functions of the ratio of the conductivity of the enclosed conductors to that of the shield. For the case of

equal conductivities of conductors and shield the optimum values are

$$\rho = \frac{c}{b} = 6.8; \qquad \sigma = \frac{h}{c} = .49. \qquad (59), (60)$$

These values may be compared with 5.4 and .46, respectively, for the pair of round conductors with circular shield. The high-frequency attenuation of each shielded quad circuit with optimum design is,

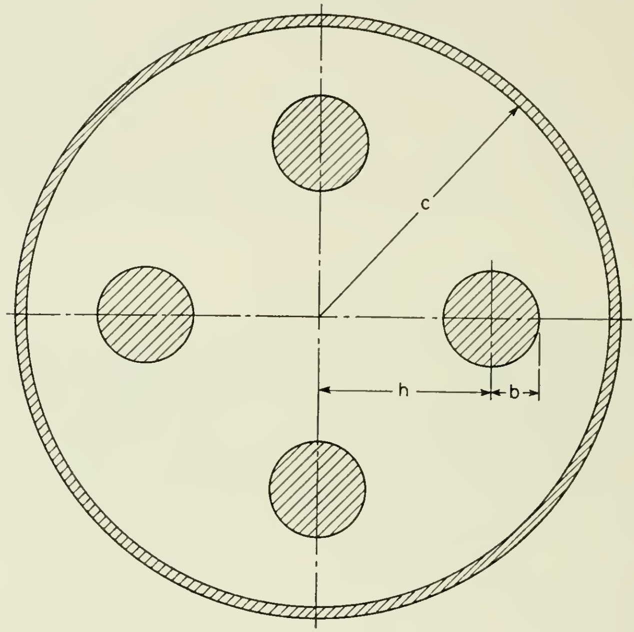


Fig. 16—Shielded quad.

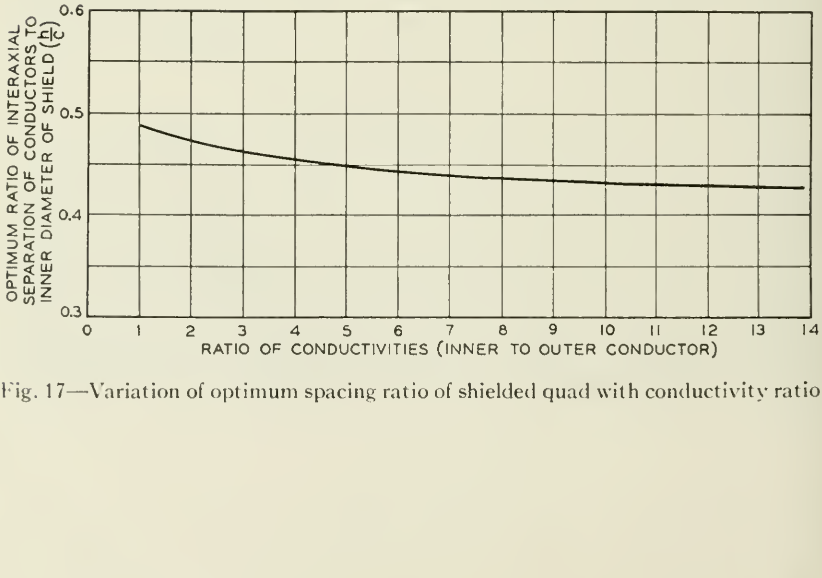


Fig. 17—Variation of optimum spacing ratio of shielded quad with conductivity ratio.

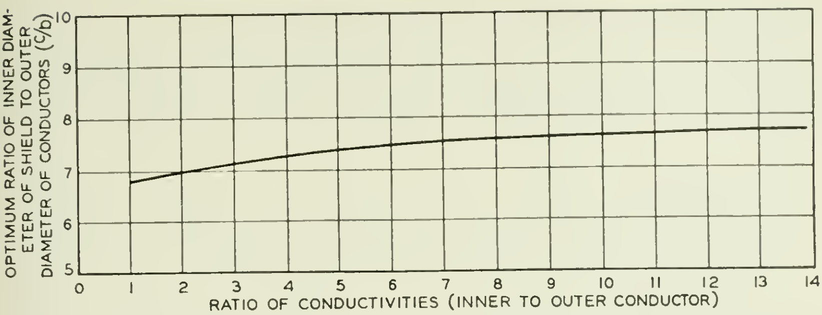


Fig. 18—Variation of optimum diameter ratio of shielded quad with conductivity ratio.

for the same diameter of shield, about 10 per cent higher than that of a shielded pair for its optimum design.

Conclusion

There have been discussed a number of different types of individually shielded circuits, both balanced and unbalanced, and the proportioning of these circuits for minimum high-frequency attenuation has been determined. The following table summarizes the optimum proportions for the more important circuits treated above. The values given are for the case where all the conductors are of the same material.

Circuit	Diameter Ratio (ρ)	Spacing Ratio (σ)
Simple coaxial	3.59	—
Double coaxial	3.59	—
Shielded pair, round conductors and circular shields	5.4	0.46
Shielded pair, round conductors and oval shield	3.7	0.47
Shielded quad	6.8	0.49

Of the transmission characteristics of these circuits, a property of particular interest is the attenuation, since, assuming adequate shielding, it is this which determines either the required repeater spacing for a given transmitted frequency band or the width of frequency band obtainable with a given repeater spacing. For each type of circuit considered there has been determined the ideal proportioning whereby the high-frequency attenuation of the circuit may be minimized. In addition a variety of methods for the solution of problems in optimum proportioning have been outlined.

It is, of course, feasible by adjustment of size to obtain the same high-frequency attenuation for all these different types of circuits. However, the size of a structure is usually reflected in its cost. An interesting picture can therefore be drawn by comparing the attenua-

tions, at the same high frequency, of different types of circuits having the same cross-sectional area and of the same material. For structures with solid wall conductors and air insulation the comparison works out as shown in the table below, the attenuation of the coaxial circuit being used as a standard of reference.

Coaxial circuit	1.00
Shielded pair, round conductors and circular shield	1.50
Double coaxial circuit	2.00
Shielded pair, round conductors and oval shield, approximately	1.3
Shielded pair, circular shield with quasi-elliptical conductors, approximately	1.47

In each case the cross-sectional area is taken as that enclosed within the shield. This neglects any differences in the thickness of shield that may be required.

A specific comparison of considerable interest is that between an unbalanced coaxial circuit and a shielded pair, the latter being taken as representative of shielded balanced circuits. The table shows that, for the same attenuation, the cross-sectional area included within the shield is larger for the shielded pair than for the coaxial circuit. On the other hand, the use of balance in addition to shielding is advantageous in that it reduces the amount of shielding needed. The shielded pair makes possible the utilization of the entire frequency range, if desired, whereas with a coaxial circuit it is necessary to discard the lower frequencies where it is uneconomical to provide adequate shielding.

A thorough-going comparison of the relative advantages and fields of application of the various types of circuits which have been discussed would extend to great length. Clearly a large number of factors enter into the choice of the configuration of shielded high-frequency circuit to be used in any given instance. These factors include the width of frequency band to be transmitted, the degree of shielding required, the relative economy of manufacture of different structures, etc. While a complete exposition of these factors has not been attempted, the principles of optimum proportioning which have been discussed should be helpful in selecting the best configuration to meet given requirements, and the particular configuration chosen should be made to conform reasonably closely to the optimum.

References

1. L. Espenschied and M. E. Strieby, "Systems for Wide-Band Transmission Over Coxial Lines," *Elec. Engg.*, Vol. 53, October, 1934, p. 1371; also *Bell System Technical Journal*, Vol. 13, October, 1934, p. 654.
2. A. B. Clark, "Wide Band Transmission Over Balanced Circuits," *Elec. Engg.*, Vol. 54, January, 1935, pp. 27–30; also *Bell System Technical Journal*, Vol. 14, January, 1935, p. 1.
3. British Patent No. 284,005, C. S. Franklin, January 17, 1928.

4. E. J. Sterba and C. B. Feldman, "Transmission Lines for Short-Wave Radio Systems," *I. R. E. Proc.*, Vol. 20, July, 1932, p. 1163; also *Bell System Technical Journal*, Vol. 11, July, 1932, p. 411.
5. S. A. Schelkunoff, "The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields," *Bell System Technical Journal*, Vol. 13, October, 1934, p. 532.
6. K. S. Johnson, "Transmission Circuits for Telephonic Communication," New York, 1925.
7. Alexander Russell, "The Effective Resistance and Inductance of a Concentric Main," *Phil. Mag.*, Vol. 17, Sixth Series, April, 1909, pp. 524–552.
8. Sir William Thomson, "On the Theory of the Electric Telegraph," *Roy. Soc. Lond. Proc.*, Vol. 7, May 3, 1855, pp. 382–399.
9. E. I. Green, U. S. Patent No. 2,029,420, February 4, 1936.
10. E. I. Green and F. A. Leibe, U. S. Patent No. 2,029,421, February 4, 1936.
11. H. A. Affel and E. I. Green, U. S. Patent No. 1,818,027, August 11, 1931.
12. S. Butterworth, "Eddy Current Losses in Cylindrical Conductors," *Royal Soc. London Phil. Trans.*, Vol. 222, May, 1922, pp. 57–100.
13. F. A. Leibe and T. C. Rogers, U. S. Patent No. 2,034,047, March 17, 1936.
14. E. I. Green, U. S. Patent No. 1,854,255, April 19, 1932.
15. E. I. Green, H. E. Curtis and S. P. Mead, U. S. Patent No. 2,034,032, March 17, 1936.
16. J. R. Carson, "Wave Propagation Over Parallel Wires: The Proximity Effect," *Phil. Mag.*, Vol. 41, April, 1921, p. 627.
17. E. I. Green and H. E. Curtis, U. S. Patent No. 2,034,033, March 17, 1936.
18. H. E. Curtis and S. P. Mead, U. S. Patent No. 2,034,026, March 17, 1936.
19. E. I. Green, U. S. Patent No. 2,034,035, March 17, 1936.
20. E. I. Green and F. A. Leibe, U. S. Patent 2,034,034, March 17, 1936.

Hyper-Frequency Wave Guides—General Considerations and Experimental Results *

By G. C. SOUTHWORTH

A peculiar form of electrical propagation is described below. It makes use of extremely high frequencies—even beyond those generally employed in radio. In some respects it resembles ordinary wire transmission but unlike the latter there are no return conductors, at least of the usual kind.

In this transmission, electromagnetic waves are sent through guides made up either of an insulator alone or of an insulator surrounded by a conductor. In a special case this insulator may be air. There are at least four different types of waves or electrical configurations that may be propagated. One of them is such that theory indicates its attenuation through a hollow conductor continuously decreases with increase of frequency. Although the paper deals largely with the nature of this transmission, some of the fundamental pieces of apparatus used in experimental work are described. They include generators, receivers and wave-meters.

Introduction

THIS paper describes a novel form of electrical propagation by means of which extremely high-frequency waves may be transmitted from one point to another, through specially constructed wave guides. The guide used for this purpose may take any one of several different forms. It may be a hollow copper pipe, which for the higher frequencies now available would be about 3 or 4 inches in diameter, or possibly a somewhat smaller conducting tube filled with some insulating material combining high dielectric constant and low loss, or it may conceivably be a rod or wire of dielectric material.[1]

The phenomena involved in this form of transmission are exceedingly interesting and at first sight paradoxical for in some cases transmission is effected through a single wire of insulating material surrounded by metal in place of a pair of metal wires surrounded by insulation. In others the wire is made entirely of insulating material. In still others electrical effects are observed only on the interior of hollow metal cavities instead of the exterior only as is ordinarily experienced. In all cases there is no return current path, at least of the kind that is commonly assumed in ordinary transmission.

The frequencies appropriate for this form of transmission begin at the higher of those generally known as ultra-high frequencies that is, 2000 mc. ($\lambda = 15$ cm.) and extend to an indefinite upper limit possibly

* To be presented at joint meeting of Amer. Phys. Soc. and I.R.E., Washington D. C., April 30, 1936.

[1] The mathematical theory of these phenomena is given in a companion paper by J. R. Carson, S. P. Mead and S. A. Schelkunoff, this issue of the *Bell System Technical Journal*.

set by the properties of available materials. These have for convenience been called hyper-frequencies. When these electromagnetic waves are propagated through either of the two forms of guide mentioned above that incorporates a metal sheath, there is little or no external field and consequently little or no interference from static or other extraneous noises.

As already mentioned there is no return conductor, at least of the kind with which we are generally familiar in ordinary wire or coaxial cable transmission. Corresponding to this difference in physical structure there are striking differences in the character of the waves propagated. On the other hand, when we compare this transmission with radio, where there might at first sight appear to be great similarity, we find little or no correspondence, for it turns out that as regards both the velocity of propagation and attenuation per unit length, radio and wave guides follow quite different laws.

In answer to the natural question as to what practical use there may be for transmission methods of this type the following considerations may be of interest: The size of structure that may be used as a guide is directly proportional to the wave-length. It happens that in structures that are at all convenient in size, the necessary frequencies correspond approximately to the highest range now being tried out in radio. If the size of structure is further reduced to make it more economical for use for long distance transmission, it is then necessary to use frequencies above this range. Thus far these can be produced and handled only with serious difficulty. Although it is possible to reduce the size of the guiding structure for a given frequency by the use of a suitable dielectric we are met with a conflicting difficulty of producing at reasonable cost the necessary medium that will incorporate high dielectric constant with sufficiently low losses. The situation then is that the art at these extreme frequencies is not yet at a point which permits a satisfactory evaluation of practical use. However, for short distance transmission or for use as antennas or projectors of radio waves or for selective elements analogous in nature to the tuning elements so commonly used in radio, there are not the same economic conditions limiting the size of structure. For such uses, then, structures of this type deserve serious consideration.

Theory indicates that one of the four types of waves (designated below as H_0) has progressively less attenuation as its frequency is increased. It happens, however, that this type requires for a given guide a higher range of frequencies than any others. This puts it, therefore, in a frequency range where the art is even less developed than for the other types of transmission and where it is even more

difficult to evaluate the economic and practical problems. This paper will, therefore, confine itself to a discussion of some of the fundamental properties of wave guides derived either from calculation or experiment. These properties include characteristic impedance, attenuation and velocity of propagation as well as frequency, selectivity and radiation.

Nature and Properties of Wave Guides

Analysis has shown that there are many kinds of waves that may be propagated through cylindrical guides. However, four of them are of unusual interest and are such as merit special consideration at this time. All four have been experimented with in our laboratory and their more important characteristics have been determined. This experimental work has been paralleled by a mathematical theory [1] to which it conforms most satisfactorily.

A good mental picture of the nature of the waves propagated through guides can probably best be had by abandoning the ordinary concept of current electricity flowing in a "go and return" circuit in favor of that of lines of electric and magnetic force. This latter concept has, of course, always been applicable even for low-frequency transmission over parallel wires or coaxial conductors but due to its complexity in pictorial representation it has usually been avoided. In the form of transmission with which we are now concerned, the field point of view is almost necessary.

Figure 1 is a pictorial representation based on this point of view of the four types of waves mentioned above as found in a guide surrounded by a metallic conductor. In these models the lines of electric force have been represented by solid lines and the lines of magnetic force have been shown by dotted lines. In the longitudinal sections, the small open circles represent lines of force directed toward the observer. The solid circles represent lines directed away from the observer. The designations E_0, E_1, H_0 and H_1 are convenient reminders of certain characteristics of these waves.

The first two waves have been designated as electric because there is a component of electric force in the direction of propagation. For similar reasons the latter have been known as magnetic waves. Such a designation is, of course, rather arbitrary and should not be construed to mean that either component resides alone. It is true here as in other forms of electromagnetic waves with which we are generally familiar, that both the electric and magnetic components are essential to the very existence of the wave and that they may conveniently be considered as different aspects of the same thing.

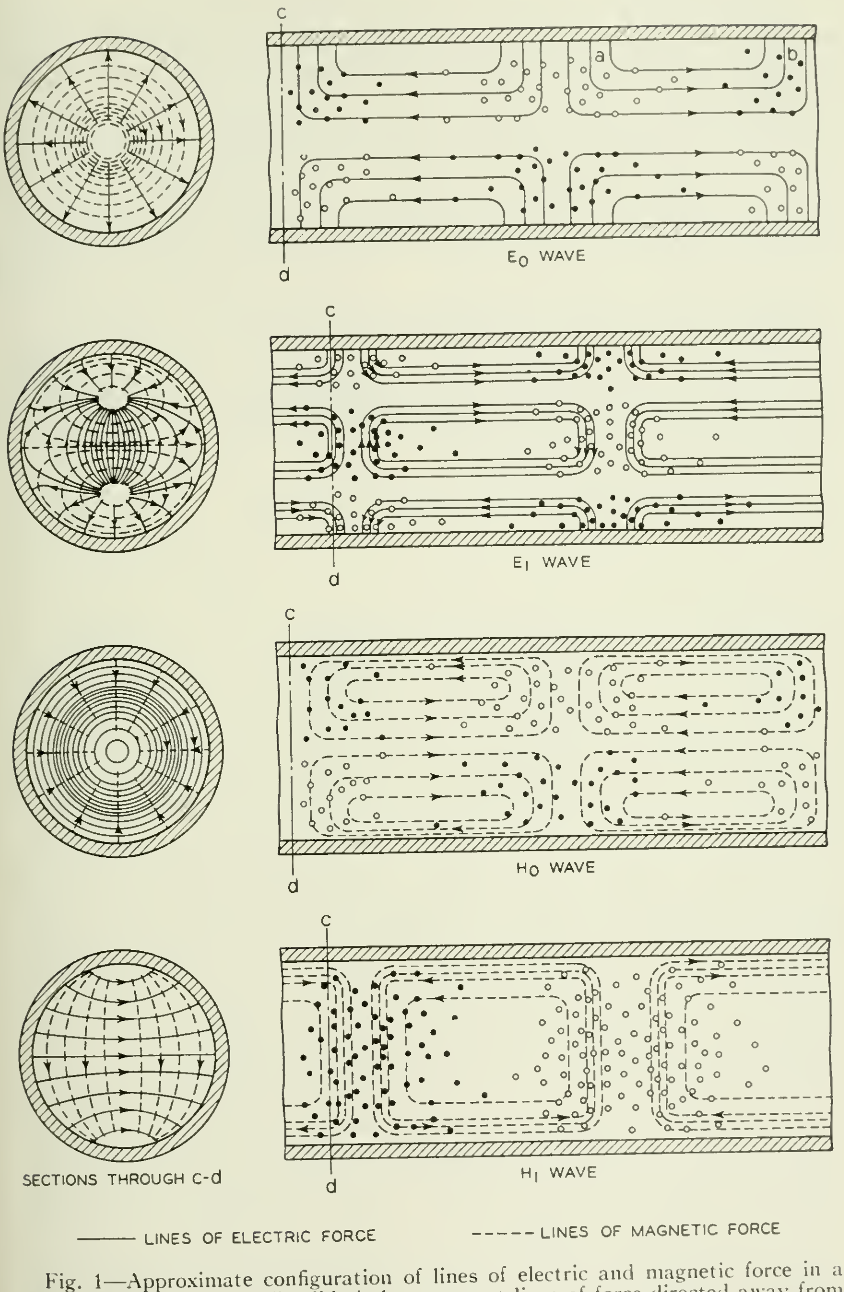


Fig. 1—Approximate configuration of lines of electric and magnetic force in a typical wave guide. Small solid circles represent lines of force directed away from observer. Propagation is assumed to be directed to the right and away from the observer.

Electromagnetic waves cannot be freely transmitted in dielectric wires or hollow conductors at all frequencies but only when the wavelength is less than a certain value set by the material of the guide and its dimensions. There is, therefore, for a given guide a critical frequency below which waves may not be propagated. We refer to this as the cut-off frequency. In a similar way we have for a given frequency, critical or cut-off diameters. These critical frequencies depend not only on the diameter (d) of the guide but on the dielectric constant (κ) of the medium as well. Also they are, in general, different for the different types of waves. For guides enclosed by a metallic conductor the cut-off wave-length is such that the circumference of the guide measured in wave-lengths is equal to the roots of certain Bessel's functions. These in turn result from solution of the Maxwell equations expressed in cylindrical coordinates. These relations are shown more fully in Table I.

TABLE I

Type of Wave	Bessel Functon	Root	Cut-off Wave-length $= \frac{\pi d \sqrt{\kappa}}{X}$
E_0	$J_0(X) = 0$	$X = 2.41$	$1.31 d\sqrt{\kappa}$
E_1	$J_1(X) = 0$	$X = 3.83$	$0.82 d\sqrt{\kappa}$
H_0	$J_0'(X) = 0$	$X = 3.83$	$0.82 d\sqrt{\kappa}$
H_1	$J_1'(X) = 0$	$X = 1.84$	$1.71 d\sqrt{\kappa}$

As a simple numerical example let us assume the H_1 type of wave having a frequency of 3000 mc. ($\lambda = 10$ cm.) being propagated in a hollow metallic pipe. The critical diameter turns out to be 5.85 cm. or roughly 2.30 inches. If the space were filled with a material having a dielectric constant of, say 5, this would have been reduced to a diameter of roughly one inch. For higher frequencies or for materials having still higher dielectric constants these critical dimensions would obviously be still further reduced and would be comparable in size to the larger conductors used in ordinary electrical practice. The critical dimensions for the other types of waves are of course larger.

Referring again to Fig. 1 we see that in the so-called E_0 type of wave, a line of electric force originating at a point a on the inner surface of the wall of the guide passes radially toward the center then axially and again radially to a corresponding point b on the inner wall roughly one-half wave-length farther along. The entire wave front as seen in cross section cut through cd consists of a symmetrical arrangement of these radial lines. The magnetic field associated with this wave consists of a series of coaxial circles shown as dotted lines not

unlike the magnetic field in a coaxial conductor such as shown in Fig. 2. The E_1 wave consists of electric and magnetic lines very similar in form to those associated with two parallel electric conductors surrounded by a metallic shield. The similarity between the fields for the two dielectric waves and the corresponding two arrangements for ordinary transmission is made more obvious by a comparison of Fig. 1 with Fig. 2. For the most part this similarity ends at this point, however, as their corresponding properties follow quite different laws.

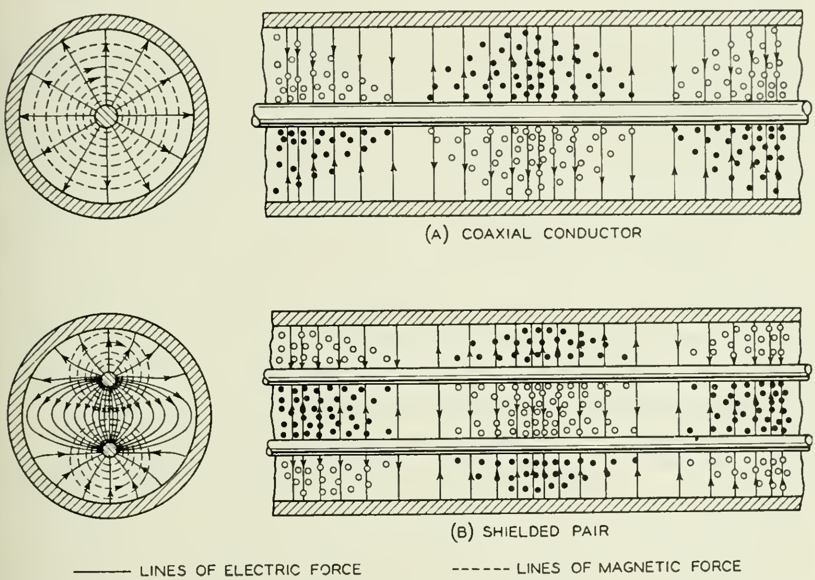


Fig. 2—Approximate configuration of lines of electric and magnetic force in a coaxial conductor and also in a shielded pair of conductors. Note similarity to E_0 and E_1 waves of Fig. 1.

The configurations of the two magnetic waves are somewhat similar to the electric waves provided we assume the electric and magnetic components to be interchanged. Nature has thus far failed to provide us with materials that possess exclusively magnetic conductivity in the sense that copper possesses electrical conductivity so there are no counterparts of Fig. 2 applicable to magnetic waves.

The general shape of the lines of electric force for all of these types of waves have been calculated. These fields have also been verified experimentally by means of a small probe consisting of a crystal detector with short pick-up wires connected to a sensitive meter. This probe was carried over the cross-section of the guide always

orienting the detector to obtain maximum deflection. These data confirmed not only the directions of the lines of force but their relative density as well.

There is one characteristic of the H_0 and H_1 configurations that at first sight seems inconsistent with our more usual views of electricity. It is the existence of a substantial tangential component of electric force apparently in close proximity to a metallic conductor. It must be borne in mind however that these frequencies are extremely high and that these distances after all represent an appreciable part of a wave-length.

If any of the four types of waves depicted above are propagated through a wire of dielectric material without the metal enclosure, lines of electric force which previously attached themselves to the inner walls of the sheath, in general, extend into the surrounding space and close as loops. This means that as the wave moves along the guide a portion of the wave power is propagated through the dielectric itself and a part through the surrounding space. The proportionate parts of the electric and magnetic fields resident inside and outside the dielectric are amenable to calculation. As might be expected they depend both on the dielectric constant of the material and on the proximity to cut-off at which the guide is operated. Results of such calculation for the E_0 type of wave are shown in Fig. 3, each for various proximities to cut-off. A dielectric constant of 81 is assumed.

For high dielectric constants and for frequencies far above the cut-off, the power is propagated largely inside the guide whereas for low dielectric constant and for frequencies just above the cut-off, a substantial amount of the power travels outside the guide. In the first case inductive disturbances communicated to neighboring guides are very small and correspondingly the guide is substantially immune to outside disturbances. In the second case these important advantages are absent.

As already stated, many of the properties of wave guides are amenable to calculation. Formulas for the purpose are included in the mathematical paper already referred to. Certain of these properties are intrinsic—as for example, velocity of propagation, attenuation and characteristic impedance. Others may be regarded as extrinsic in that they result largely from the manner in which the guide is used. Examples of the latter are frequency-selectivity and radiation.

Velocity of Propagation

It will be remembered that the velocity of electric waves over ordinary conductors immersed in a particular medium is substantially that of light for that medium. In other words it is equal to the velo-

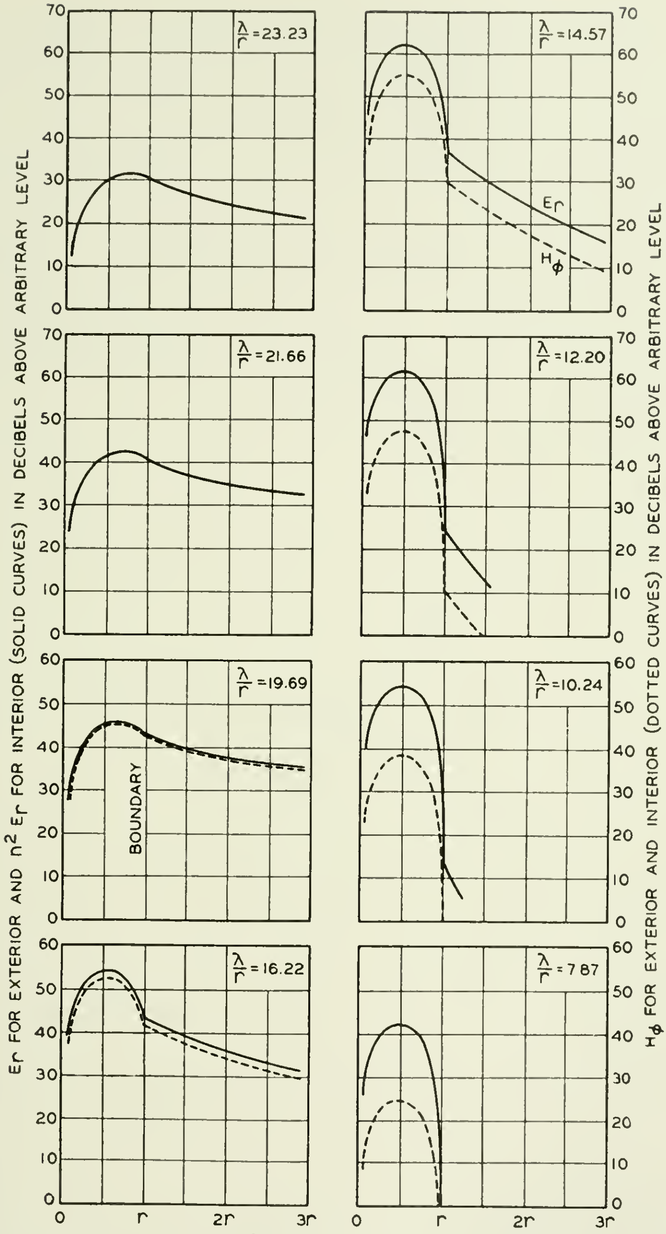


Fig. 3—Relative intensities of electric and magnetic fields inside and outside a dielectric wire while propagating the E_0 type of wave. A dielectric constant of 81 is assumed.

city of light in free space divided by the index of refraction (square root of the dielectric constant). It is also dependent to a small extent on the resistance and permeability of the conductors themselves. The velocity of propagation in wave guides depends not only on these properties but also on the dimensions of the guide as well. For a cylindrical guide it is convenient to express the relation between frequency and dimension as a ratio of wave-length in free space to diameter (λ/d). Also the velocity in the guide may conveniently be expressed as its ratio to the velocity of light ($c/v = k$). Designating by λ the wave-length in free space and by λ_g the wave-length in the guide $k = \lambda/\lambda_g$. Figures 4, 5 and 6 show in graphical form these velocity ratios for three representative cases. The solid curves are calculated. The points are experimental.

Figure 4 covers the case of E_0 waves in a dielectric having a constant

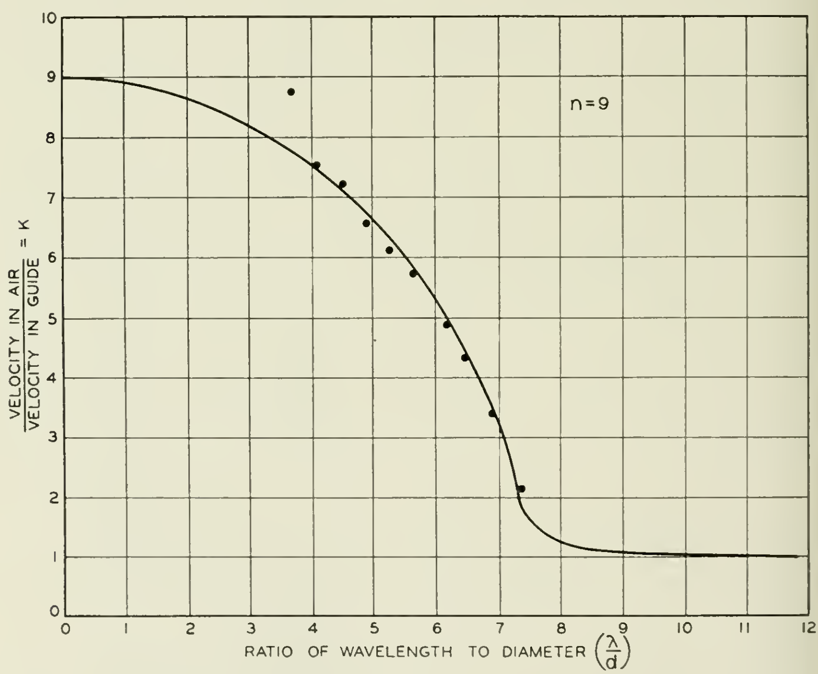


Fig. 4—Velocity ratio for the E_0 type of wave in a dielectric wire ($K = 81$).

of 81 when surrounded by air. It will be observed that at the highest frequencies (lowest values of λ/d) the velocity of propagation is one ninth that of light in free space whereas at the lower frequencies (near cut-off) the velocity is that of light in free space. If the di-

electric constant were progressively lowered the velocity even at the highest frequencies would approach that of light in free space and the curve shown would become progressively flatter. In the limit the dielectric constant would be unity and the velocity ratio also would be unity. Under this circumstance the dielectric wire having a constant substantially the same as that of the surrounding medium would cease to function as a guide.

The experimental points of Fig. 4 were obtained by transmitting waves at each of several frequencies ranging from 100 mc. (λ = 300 cm.) to 400 mc. (λ = 75 cm.) through columns of moderately pure water. The distances between nodes and loops of the standing waves gave data for the velocity of propagation. The method therefore utilized, in a modified form, a technic sometimes invoked for determining the velocity of electric waves on wires or the velocity of sound in air columns. The columns were supported in thin walled bakelite cylinders each about three feet long. Two diameters were used, 6 inches and 10 inches respectively.

Figure 5 covers the case of the same type of waves and the same

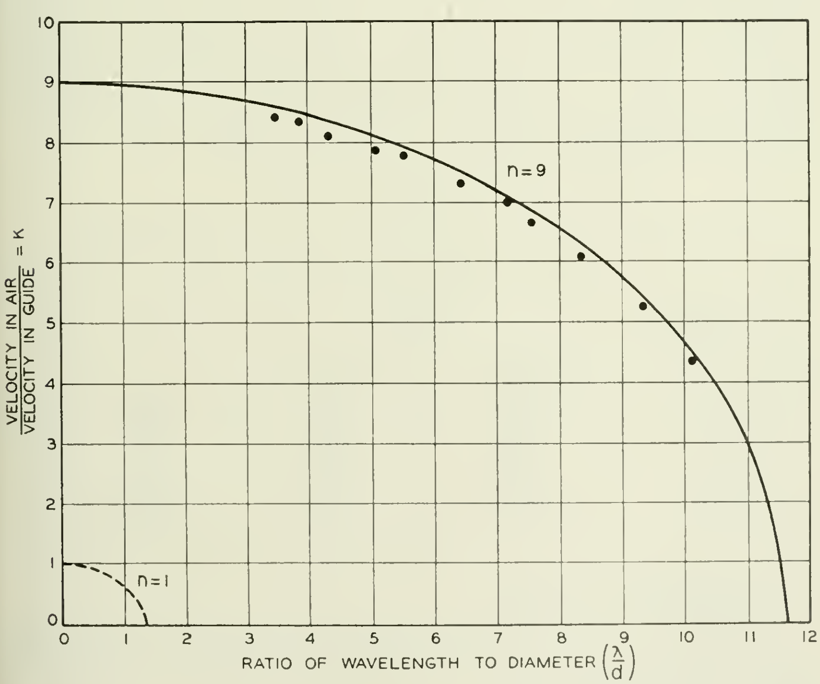


Fig. 5—Velocity ratio for the E_0 type of wave in a metal pipe filled with an insulator (K = 81).

dielectric as above but surrounded by metal. It will be noted that the limits of this curve are essentially the same as for the unshielded guide but that the two curves follow rather different courses. If, in this case, the dielectric constant were progressively reduced the curve shown would gradually shrink into the miniature replica shown dotted in the lower left corner. This is of course the practical case of a hollow conductor to be discussed shortly. The above discussion leads naturally to the view that a wave guide is a propagating medium bounded by a dielectric discontinuity. In one case the discontinuity is the interface between the dielectric of the guide and the surrounding medium. In the other it is the interface between the dielectric and a surrounding conductor.

It will be noted from Fig. 5 that at the highest frequencies the phase velocity in shielded guides, like that for unshielded guides, is the same as the velocity along ordinary conductors in that medium but at frequencies near the cut-off this velocity approaches infinity. The solid curve is calculated on the assumption that the medium had a dielectric constant of 81. The indicated points are the results of experiments made with water as a dielectric. For this experiment the water was supported in three-foot cylinders of copper, six inches and ten inches in diameter respectively. The same range of frequencies was used as above. A somewhat closer agreement between calculation and experiment would have resulted if a value of dielectric constant of 78.9 had been assumed in the computations.

TABLE II

VELOCITY RATIOS FOR H_1 WAVES IN HOLLOW CONDUCTORS

Ratio Space Wave-length to Guide Diameter [1]	Ratio Velocity in Free Space to Velocity in Guide		Difference	Per Cent
	Calculated [3]	Measured [2]		
0.980	0.818	0.818	0.000	0.0
1.033	0.795	0.797	0.002	0.3
1.108	0.757	0.762	0.005	0.7
1.246	0.684	0.683	− 0.001	− 0.1
1.375	0.592	0.601	0.009	1.5
1.469	0.510	0.514	0.004	0.8
1.547	0.424	0.429	0.005	1.2

[1] Probable error 0.4 per cent.
[2] Probable error 0.4 per cent.
[3] Probable errors (arising from error in λ/d) range from 0.2 per cent for $\lambda/d = 0.98$ to 1.8 per cent for $\lambda/d = 1.55$. Note that agreement in most cases is within probable error. However the fact that in all but one observation differences have same sign suggests some systematic relation.

Figure 6 is based on calculations covering the case of a hollow conductor (air dielectric) propagating the H_1 type of wave. The experimental data for plotting points on Fig. 6 were obtained at frequencies extending from 1500 mc. (λ = 20 cm.) to 2000 mc. (λ = 15 cm.) on hollow cylinders ranging in diameter from four inches to six inches. Relative velocity was determined from the length of standing waves set up in short sections of these wave guides. The measurements represented were made with much more refined apparatus than utilized in obtaining the data for Figs. 4 and 5.

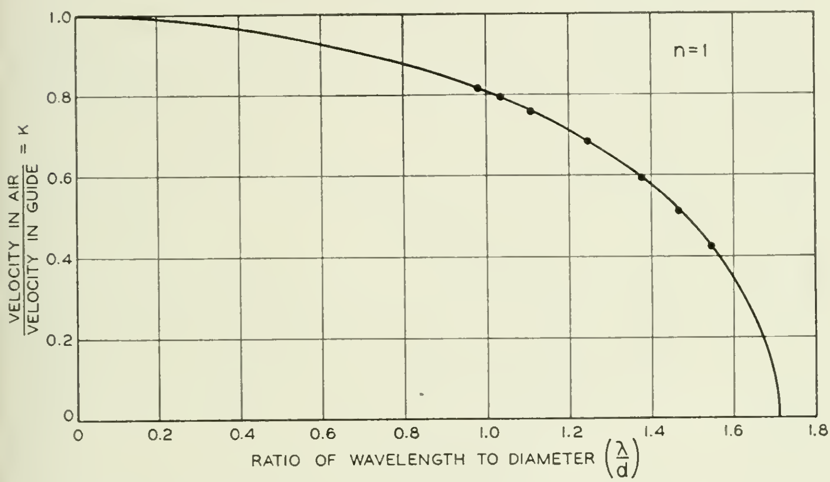


Fig. 6—Velocity ratio for the H_1 type of wave in a hollow metal pipe.

Attenuation

Figure 7 shows in graphical form the calculated attenuations suffered by each of the four more common types of waves when traveling through a hollow copper pipe 5 inches in diameter. It is immediately obvious that the attenuation is infinite for all waves at their respective cut-off frequencies. However, at frequencies above the cut-off this attenuation becomes finite, generally descending to values comparable with attenuations experienced on ordinary conductors at considerably lower frequencies. For the E_0 and E_1 types of waves the attenuation falls from infinity at cut-off to a minimum at a frequency $\sqrt{3}$ times the cut-off frequency after which it again begins to increase and ultimately varies in a linear fashion much as does attenuation over ordinary conductors. For the H_1 type of wave this minimum comes at a frequency 3.15 $\sqrt{3}$ times the cut-off frequency. Thus we see that

for these three types, wave guides are somewhat similar in their behavior to ordinary conductors when operated at the highest frequencies but they depart radically at frequencies near cut-off.

Calculations indicate that the H_0 type of wave has a descending attenuation characteristic at all frequencies above cut-off. This suggests that we may be able to realize very low attenuation merely

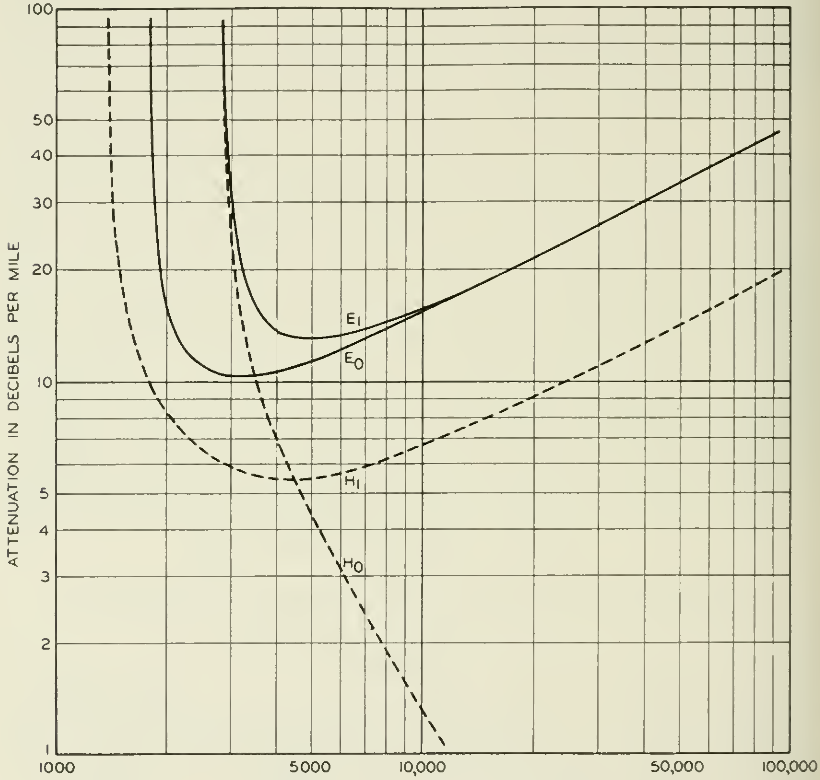


Fig. 7—Attenuations suffered by each of the more common types of waves in a hollow copper pipe 5 inches in diameter.

by raising frequency. This remarkable property is, so far as the author is aware, altogether unique in the realm of electrical transmission. It should be borne in mind, however, that for structures having reasonable dimensions, these low attenuations can only be obtained from frequencies that are above those now readily available. It may be noted in passing that at the minimum of the H_1 curve, transmission is flat to a half db per mile over a band-width of 4000 mc.

The author's experimental work on attenuation is still incomplete, but the results to date are altogether in keeping with calculation. Work done at or near cut-off for all four types of waves confirms their descending characteristics at these points. Other more systematic measurements made on the H_1 type of wave over a considerable range of frequencies are also in good agreement with calculation. Typical results are shown in Fig. 8. They were made on a straight section of

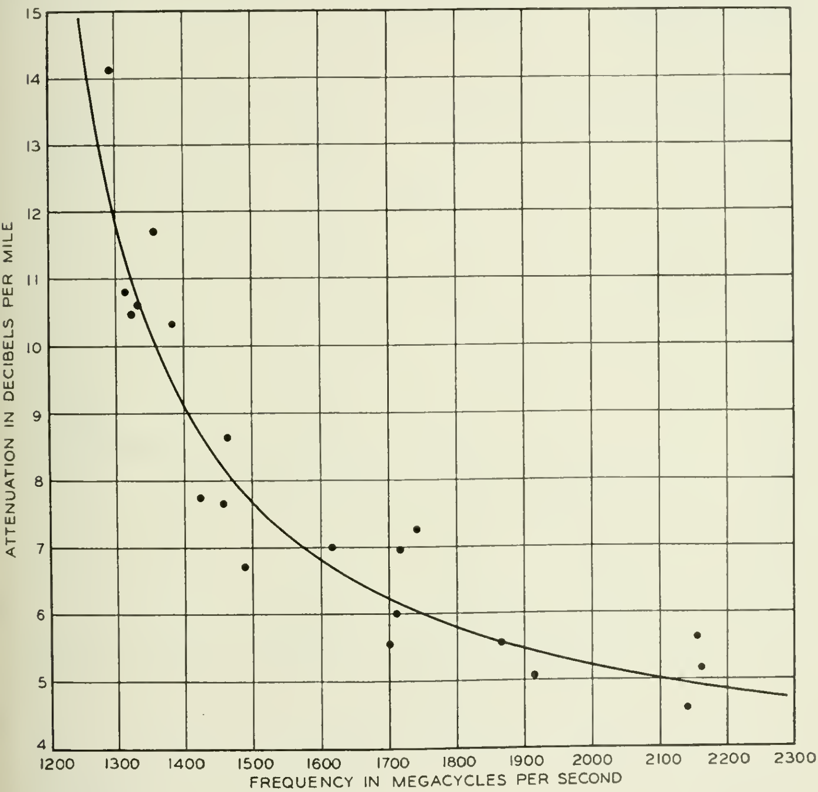


Fig. 8—Attenuation suffered by H_1 waves in a 6-inch hollow copper pipe. Curve is calculated. Plotted points are experimental.

hollow copper pipe six inches in diameter and 1250 feet long. No experimental attenuation data on the H_0 type of wave are yet available except at cut-off. It may be argued, however, that the same theory applies to all four forms of waves so that data tending to confirm the calculated attenuation of one form of wave tends also to substantiate the predicted attenuation for the other forms as well.

Characteristic Impedance

A second intrinsic property of wave guides is characteristic impedance. It may be calculated by integrating the complex Poynting vector over the cross-sectional area and dividing the result by the square of the effective current. Formulas for this purpose are included in the companion mathematical paper referred to above. The numerical results of such calculation are shown in Fig. 9 for a 4-inch diameter hollow copper conductor for each of the four principal waves mentioned above.

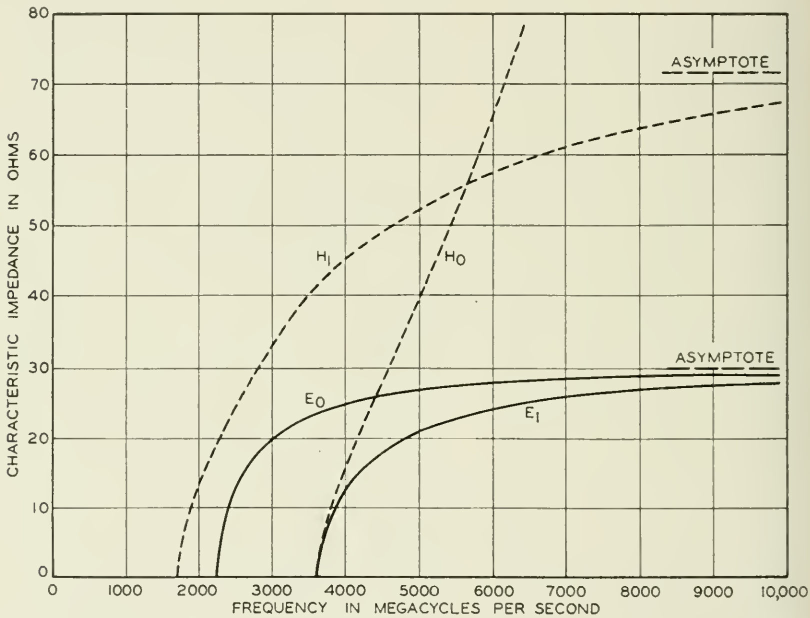


Fig. 9—Calculated values of characteristic impedance of a 4-inch hollow copper pipe for each of the four more common forms of waves.

It will be remembered that when an ordinary wire line is terminated in its own characteristic impedance there are no standing waves. This condition leads to a maximum of power delivered to the receiver. Such an impedance match is sometimes referred to simply as a termination. Terminations for wave guides are entirely similar in their behavior to those of wire lines and may be had by a variety of means. One is a thin film of resistance material placed perpendicular to the axis of the guide followed at a prescribed distance by a perfectly conducting reflector. It is often convenient to provide the latter in

the form of a movable piston. Another form is a resonant chamber containing some dissipative material. Conditions for termination may be calculated in so far as the properties of materials are known or they may be determined experimentally by successive adjustments of film density and piston adjustments until standing waves have been eliminated. Figure 10 shows graphically a typical series of experi-

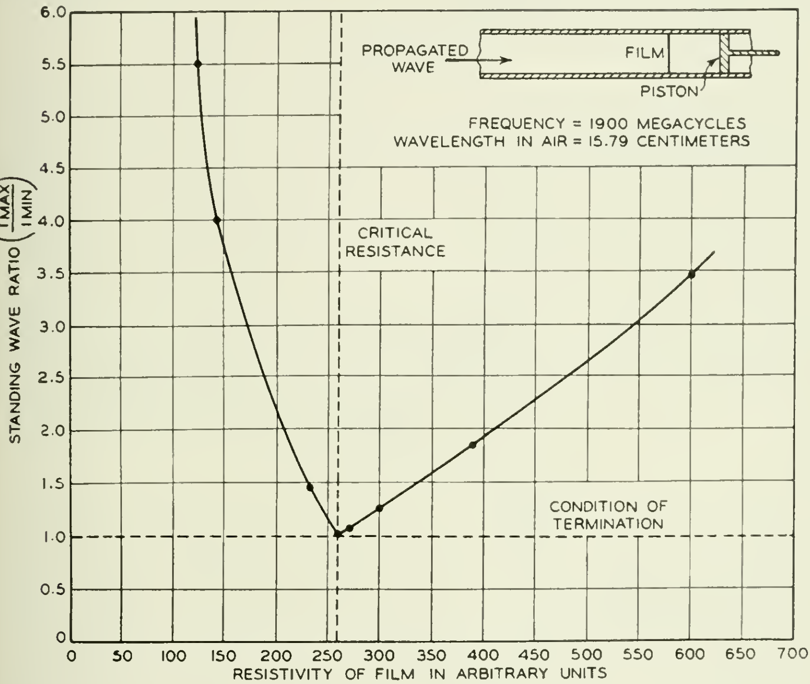


Fig. 10—Typical set of experimental data as various degrees of impedance match are obtained.

mental data of the magnitudes of standing waves as various degrees of impedance match are obtained.

Frequency Selectivity

It is evident from Fig. 7 above that wave guides are inherently high-pass filters. There is still another property of a wave guide that may also provide selectivity. It depends on the principle of standing waves. By this means, resonance effects may be produced that make a short section of guide behave somewhat as if it were a simple series circuit consisting of an inductance and a capacity in series

with an electromotive force. Under other conditions, it may behave as a circuit made up of inductance and capacity in parallel with an electromotive force.[2] At still other frequencies it may present to a source a positive (inductive) reactance or a negative (capacitive) reactance. This makes possible circuit elements which may be combined to form various filter or network equivalents. We may have, therefore, from wave guides frequency selection by either or both of two fundamentally different properties.

Radiation

Discontinuities in wave guides, particularly those in which no shield is present, tend toward losses by radiation. In the case of a hollow conducting pipe radiation issues from the open end much the same as sound waves from a hollow tube. It has been possible to expand the ends of these pipes into horns, thereby obtaining effects very similar to those common in acoustics. Such an electrical horn not only possesses considerable directivity but it may also provide a moderately good termination for the pipe to which it is connected. In so doing its function is probably quite analogous to that of a true acoustic horn which provides an efficient radiating load for its sound motor.

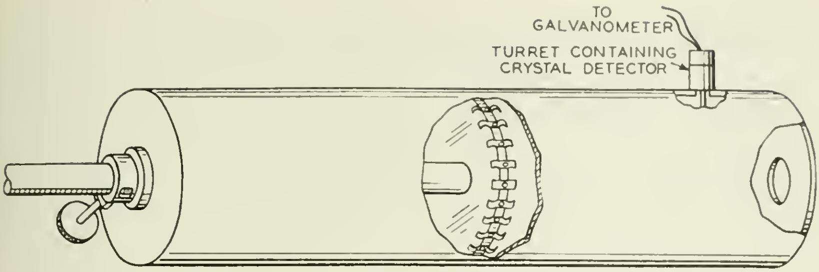
Some Apparatus and Methods Used in Wave Guide Studies

It is obvious, of course, from the very nature of guided waves that the apparatus and methods must be rather different from the more common electrical methods. This difference is such that an adequate description would require more space than is here available. However, for purposes of completeness a few of the more interesting and fundamental aspects of the experimental side are included below. For the most part this description will center around the H_1 type of wave. (See Fig. 1 above.)

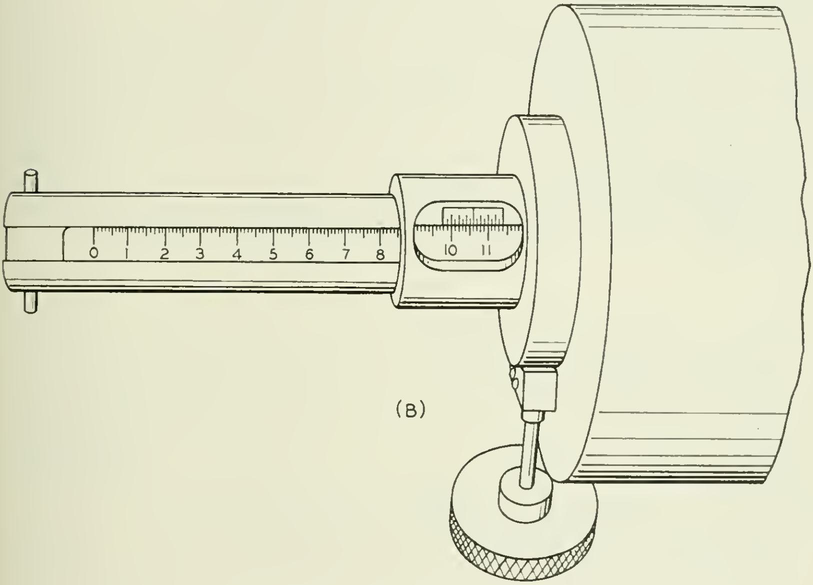
The Simple Resonant Chamber

In much the same way that the simple tuned circuit containing localized inductance and capacity is fundamental to the radio art so also is the simple resonant cavity fundamental to wave guide work. Although it may assume a variety of forms, one of the more obvious is a short piece of cylindrical wave guide, preferably of hollow metal pipe bounded by a piston and an iris diaphragm as shown in Fig. 11.

[2] In pursuing this work it has been convenient at times to refer not only to *circuit* analogues but also to *optical* and also *acoustical* analogues. This has been due in part to the lack as yet of an adequate vocabulary and in part to the hybrid nature of the subject at hand.



(A)



(B)

Fig. 11—A tunable chamber resonant to short electric waves.

In its role as a tuned circuit, the resonant chamber is sometimes used as a wave-meter, sometimes in connection with a generator of short waves (thereby enabling a vacum tube to work more effectively) and sometimes as an element in a receiver (thereby impressing on a detector a maximum of the received wave power). When such a chamber is excited by very short electric waves and is varied in length, resonance takes place at certain specified intervals depending on the frequency and phase velocity. This condition may be detected either by a crystal detector and meter located just outside of the iris opening or by a somewhat more elaborate arrangement whereby a crystal

mounted in a shielded cartridge is coupled to the chamber by a probe wire perhaps a quarter inch long extending through a hole in the wall. Fig. 11*A* shows one form of resonant chamber complete with detector. The piston position is read off on a scale and vernier (Fig. 11*B*). Successive positions at which resonance is noted give data for determining velocity of propagation. Chambers of this kind having various diameters were used to verify the velocity ratios shown earlier in this paper. Electrical connection between piston and walls may be had by numerous phosphor bronze fingers, or perhaps by ball bearings distributed in a race around the periphery. Good contact is not always essential. In fact, fair work may sometimes be done with a loosely fitting piston or even an insulated piston.

Resonant chambers may be activated merely by placing them within a foot or two of a source of waves such as a Barkhausen oscillator. Their dimensions must, of course, conform to the wave-length requirements as outlined above. Standard 5-inch OD brass pipe having one-sixteenth inch wall has been found satisfactory for the frequency range from 1500 mc. (λ = 20 cm.) to 2000 mc. (λ = 15 cm.). Any convenient length around 2 feet is appropriate for the variable type of chamber.

Generators

One arrangement for generating the H_1 type of wave consists of connecting the primary source of waves between diametrically opposite points on the inside of a hollow cylindrical conductor as shown by Fig. 12*A*. This primary source may consist of a positive grid (Barkhausen) tube or a magnetron.[3] Both have been used successfully to give frequencies up to about 3330 mc. (λ = 9 cm.).

A typical arrangement of such an oscillator is shown in Fig. 12*B*. The terminals of the spiral grid of the Barkhausen tube are connected to diametrically opposite points through a suitable by-pass condenser. The filament and plate leads enter along a plane perpendicular to that of the grid. Since the grid leads correspond to lines of electric force in the generated wave, the diametral plane perpendicular thereto corresponds to an equipotential. By locating the plate and filament leads in such an equipotential, their presence will not materially affect the normal field prevailing in the chamber. In the design shown the filament connectors constitute the outside plates of a three-plate by-pass condenser. The third or central plate is a rigid member grounded on the main guide. It connects to the plate of the Barkhausen tube. Connections to the exterior are had through five

[3] "Vacuum Tubes as High-Frequency Oscillators," M. J. Kelly and A. L. Samuel, *B.S.T.J.*, Vol. 14, p. 97, January 1935.

insulated binding posts. The oscillator unit shown carries on its exterior a plug connector leading by cable to a nearby d-c. power supply unit.

If an oscillator similar to that described above were connected into the middle of a long hollow pipe, waves, would of course, be propagated in both directions. Those that would ordinarily be propagated to the left may be reflected by a suitably located reflecting wall or piston so

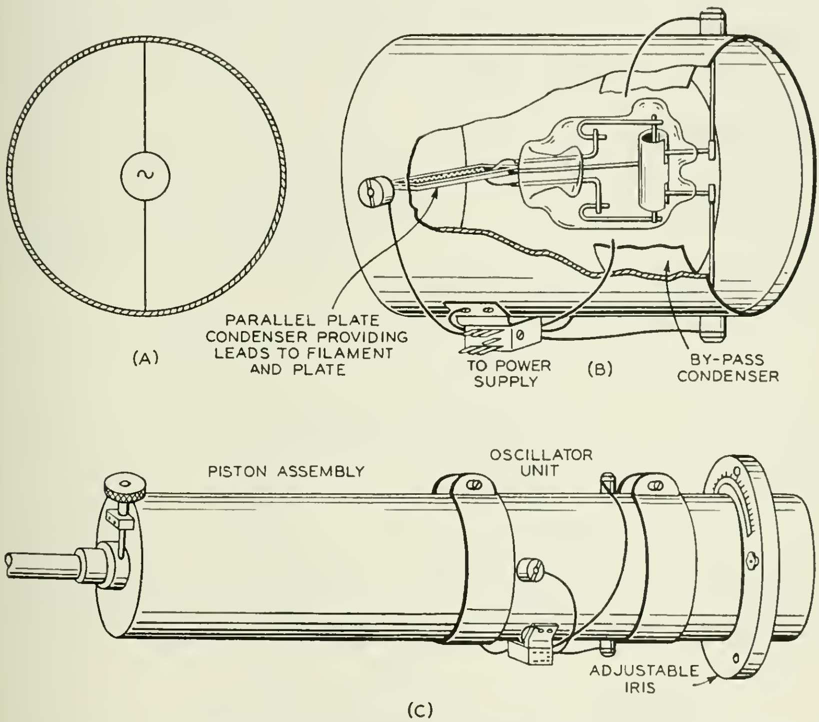


Fig. 12—Various component parts of a wave guide generator. (*A*) Schematic representation. (*B*) The oscillator unit. (*C*) Complete generator including oscillator piston and iris.

as to reinforce those being propagated to the right. Also an iris of suitable proportions may be so located in front of the generator as to further enhance oscillations. As has been pointed out above the section of pipe bounded by the piston and iris together approximate in behavior a tuned circuit. It is convenient to regard the chamber as a load impedance characteristic of the tube itself or perhaps it should more properly be regarded as a transformer by which the oscillator is matched to the line.

In practice the generator may conveniently be built up from an oscillator unit, a piston assembly and an adjustable iris, all of the same diameter of pipe fastened together by exterior metal clamps as shown in Fig. 12*C*. The open end of this generator may be connected to a guide over which transmission is desired or it may be coupled loosely to some nearby laboratory apparatus on which measurements are to be made.

The total length of the chamber and hence the piston setting will of course depend on the frequency to be generated. In general this will be roughly an integral number of half wave-lengths. The relative position of the oscillator along the length of the chamber will depend on its impedance characteristics and to some extent on the diameter of the iris opening. For a piece of laboratory apparatus where frequency variability is desired these various dimensions should preferably be adjustable as shown. If a source of single frequency is desired, the resulting apparatus may be greatly simplified as all of these dimensions may be fixed at the time of construction.

The Tuned Receiver

By reversing the principle used in the generator above, replacing the oscillatory source by a suitable indicator the resonant chamber becomes effectively a simple tuned receiver. If the indicator is appropriately located along the length of the chamber, substantially all of the incident power will be absorbed and the device as a whole will be a veritable sink of wave power. It may be clamped to the end of a long wave guide, thereby constituting a termination, or it may be used to pick up short radio waves of not too small amplitude. See Fig. 13.

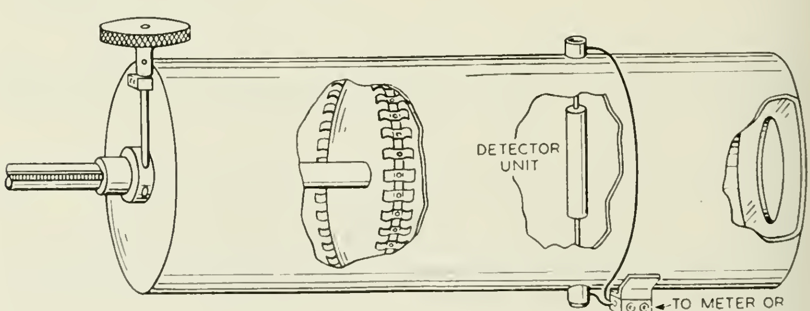


Fig. 13—A tuned receiver based on the resonant cavity principle.

Indicators

It is often desirable to have available in the laboratory some kind of a wave indicator or probe such as shown in Fig. 14. This one consists of a simple silicon detector in cartridge form, together with a microammeter, both mounted on a fibre support of convenient size and shape for exploring the fields prevailing around any piece of apparatus. It is easy to show by this means that there are no appreciable fields prevailing around a generator such as described above except near the orifice. Also this probe may be used to determine the

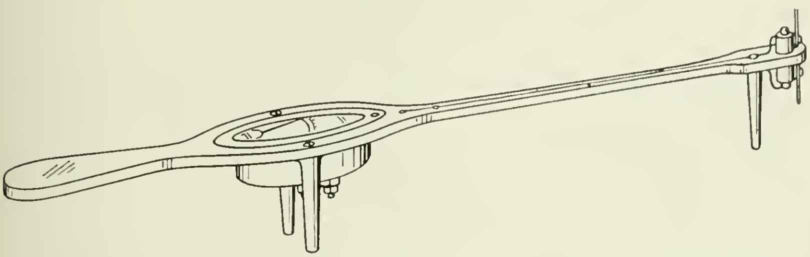


Fig. 14—A convenient probe for exploring the field around a source of waves.

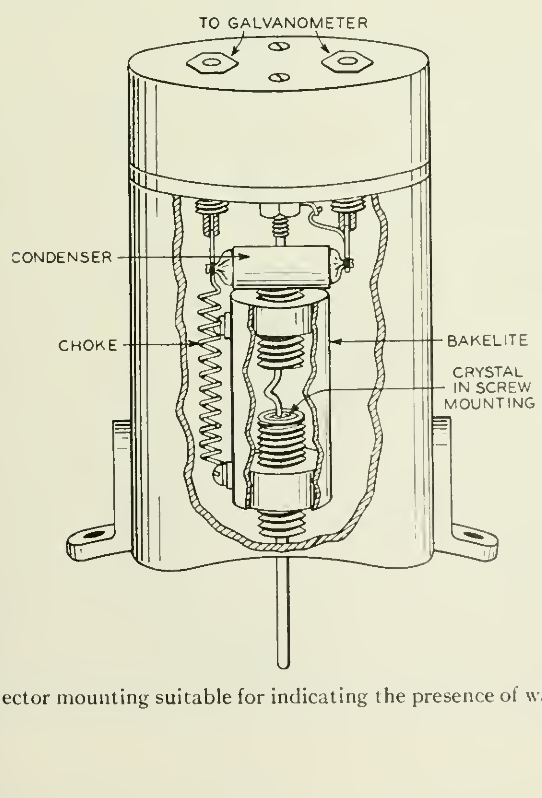


Fig. 15—A detector mounting suitable for indicating the presence of waves in a guide.

approximate orientation of the lines of electric force in the wave front as well as the general directive pattern of the radiation.

Figure 15 shows details of a crystal detector mounting suitable for an indicator on a wave-meter. When silicon crystals are used, units may be had that will hold their calibrations moderately well over considerable periods of time. Thermocouples of both the cross wire and deposited type have been used with moderate success. Also diode and triode rectifiers have been tried. However, for general laboratory use where simplicity and convenience are important the crystal detector is perhaps best.

Wave-Meters

It is, of course, desirable in this work to know the frequency or wave-length being used. The simple resonant chamber already described enables wave-length to be measured accurately. However, such a device does not give directly the wave-length in free space since in these chambers phase velocities are in general greater than ordinary light. It is true, of course, that a suitable conversion curve can be prepared. However, it is often more convenient to use for a wave-meter some form of a coaxial conductor system on which standing waves may be measured. These will be very nearly at least the length of the corresponding waves in free space.

Figure 16 shows a wave-meter based on this principle. The conductor (a) and the hollow cylinder(b) constitute the coaxial conductors. A bridge (c) in the form of a conducting disc is made movable by means of the threaded tube (d) which passes over the central conductor (a). This tube carries a millimeter thread engaged by the knurled head (e). One complete turn of this head therefore raises or lowers the bridge by one millimeter. If coarse adjustments are desired the head may be disengaged from the threaded tube by a cam operated by the knob (f), and the tube raised or lowered by taking hold of its extended portion. The outer conductor or shell carries an open slot (g) through which an index (h) attached to the shorting bridge (c) extends. This index passes over a centimeter scale (i). The outer conductor is mounted on a short piece of wave guide so that the apparatus may be clamped in line with other apparatus. The inner conductor (a) extends through a small opening in the section of wave guide far enough to extract from the passing waves enough power for activating the wave-meter. This coupling may be varied as needed by extending or retracting a third small rod (jj') running through the center of central conductor (a).

Resonance is indicated by a crystal detector (k) and d-c. meter. This detector is only loosely coupled to the coaxial system by a small

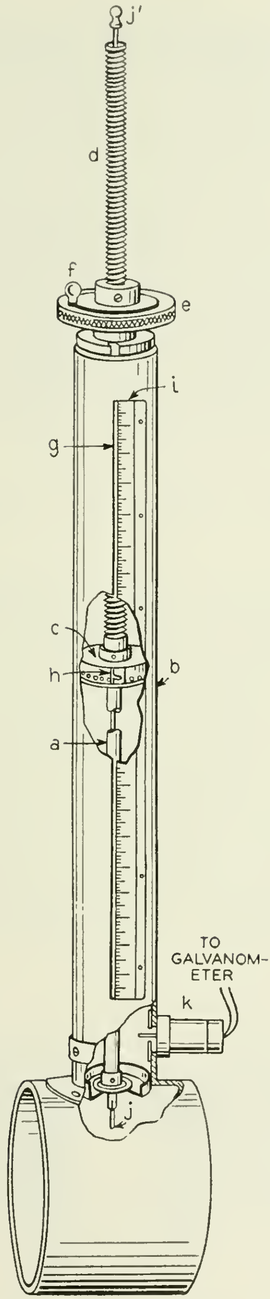


Fig. 16—A form of coaxial conductor wave-meter particularly adaptable to wave guide work.

pick-up wire extending through the walls of the hollow cylinder. This form of wave-meter is moderately fast and permits wave-length differences of one or two hundredths of a centimeter to be readily detected.

Miscellaneous Apparatus

Sometimes it is desirable to change the length of a pipe without changing its diameter. For this purpose telescoping pipe is to be avoided. A pipe with removable sections may, however, be provided. Units of 10 cm., 5 cm., 2 cm., 2 cm. and 1 cm. have been found convenient. They are aligned in a slightly larger half-section of the same kind of pipe which provides their support.

It may be desirable at times to investigate the field inside a pipe to determine if standing waves are present. This may be done by mounting a detector similar to that shown in Fig. 15, on a carriage so it may be advanced along a slot cut in a piece of wave guide perhaps 60 cm. long. Often it is necessary to pass from one size of pipe to another. A conical reducer perhaps 30 cm. long may be used for this purpose.

It is usually desirable to construct components such as the above with lengths of some integral number of centimeters such as 10 cm., 20 cm. or 50 cm. This obviously facilitates the addition of the component lengths used and often simplifies calculation.

It is obvious from the above that a laboratory working with wave guides must use for its circuit components such unusual electrical items as hollow pipes, movable pistons and iris diaphragms. These should be capable of quick assembly into a variety of forms, sometimes as a generator, sometimes as a tuned receiver and sometimes as a termination. This object imposes a wide range of requirements that can best be met by mounting the parts by means of clamp supports on a saw-horse arrangement or wave guide bench such as shown in Fig. 17.

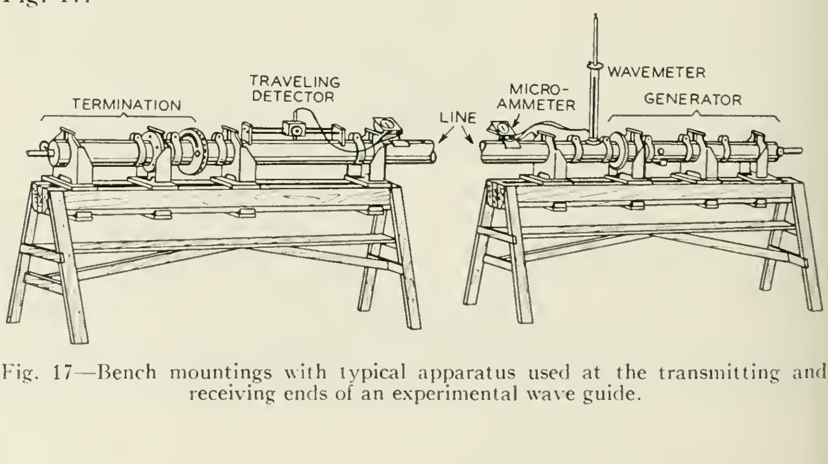


Fig. 17—Bench mountings with typical apparatus used at the transmitting and receiving ends of an experimental wave guide.

REFERENCES

1. "On the Passage of Electric Wave Through Tubes, or the Vibrations of Dielectric Cylinders," Lord Rayleigh, *Phil. Mag.*, Vol. 43, 1897, pp. 125–132.
2. "Electromagnetic Waves in Dielectric Wires," Hondros and Debye, *Ann. d. Phys.*, Vol. 32, 1910.
3. "Detection of Electromagnetic Waves on Dielectric Wires," Zahn, *Ann. d. Phys.*, Vol. 49, 1916.
4. "Electromagnetic Waves in Dielectric Conductors," Schriever, *Ann. d. Phys.*, Vol. 63, 1920.
5. French Patent 757,972, October 23, 1933.
6. British Patent 420,447, December 3, 1934.
7. "Measurements in the Radiation Field of a Linear Antenna Excited Inside a Hollow Cylinder," Bergmann and Krügel, *Annalen der Physik*, Vol. 21, No. 2, page 113, October 29, 1934.
8. A Paper entitled "Transmission of Electromagnetic Waves Through Hollow Tubes of Metal," by W. L. Barrow of Massachusetts Institute of Technology, is in an advanced stage of preparation. It is intended for presentation before the Washington Meeting of the U.R.S.I. and I.R.E., May 1, 1936, and for publication in an early issue of the *Proceedings of the Institute of Radio Engineers.*

Hyper-Frequency Wave Guides—Mathematical Theory

By JOHN R. CARSON, SALLIE P. MEAD and S. A. SCHELKUNOFF

Following a brief historical sketch, this paper deals with the mathematical theory of wave transmission in two novel kinds of cylindrical wave guides of circular cross section; namely, the hollow conductor and the dielectric wire. These transmission systems behave as high pass filters with exceedingly high critical frequencies.

The attenuation and impedance characteristics of the hollow conductor, heretofore ignored as far as the writers are aware, are given especial attention. This investigation discloses the remarkable fact that there exists in this system *one* and *only one* type of wave, the attenuation of which *decreases* with increasing frequency, a characteristic which attaches to no other type of guided wave known to the writers.

I. Introduction

THE object of this paper is to derive and discuss the characteristics of two novel guided wave *transmission* systems. Structurally one consists simply of a straight hollow [1] conducting cylinder of circular cross-section. The electromagnetic wave is confined inside the cylindrical sheath and is propagated along the axis of the cylinder. The other consists simply of a dielectric wire, within which the major part of the electric field is confined. The mathematical theory developed below does not deal with the question as to how such waves are established nor with the reflection phenomena which must occur at the terminals and other points of discontinuity. The analysis is limited to finding the types of waves which are possible in such systems, and to investigating and describing their characteristics.

The historical background of the problem is interesting. In 1897 Rayleigh published a paper entitled "On the Passage of Electric Waves through Tubes, or the Vibrations of Dielectric Cylinders." [2] Dealing solely with ideal cylinders of perfect conductivity he showed that for all types of waves that can exist inside the cylinders there are critical frequencies below which the waves are attenuated and above which they are freely transmitted. The first paper on transmission along dielectric wires was that published in 1910 by Hondros and Debye entitled "Elektromagnetische Wellen an dielektrischen Drähten." [3] This deals theoretically with transmission along cylinders of ideally non-conducting material, somewhat along the lines followed in Section IV

[1] The term *hollow* means that the interior of the cylinder is electrically non-conducting.

[2] *Phil. Mag.*, Vol. 43, 1897, pp. 125–132.

[3] *Ann. der Phys.*, Vol 32, 1910, pp. 465–476.

of this paper. Another paper, entitled "Über den Nachweis elektromagnetischer Wellen an dielektrischen Drähten,"[4] published in 1916 by Zahn, is of interest because of the historical note attached, which indicates that experimental work was begun in 1914 by Rüter and Schriever, two students of Zahn, and continued with such diligence as the exigencies of war permitted until the date of Zahn's paper, at least. In 1920 Southworth, then working at Yale University, accidentally observed such waves in a trough of water which he was using in connection with some high-frequency studies, measured their wave-lengths and recognized their identity with those discussed by Schriever[5] in a paper which appeared at about that time. In 1924 Carson rediscovered the transmission characteristics of the hollow conducting cylinder, and disclosed it in an unpublished memorandum entitled "Hyper-Frequency Wave Filters." Finally, in 1931, Southworth, then a research engineer with the American Telephone and Telegraph Company, returned to the subject and initiated the comprehensive investigation which he is reporting in a companion paper.[6] Independently, and almost simultaneously, Hartley, at the Bell Telephone Laboratories, suggested the possibility of guided transmission along a *hollow* cylindrical dielectric wire; and these two (Southworth and Hartley) enlisted our cooperation in a mathematical investigation.

In the theoretical parts of these papers dissipation was always neglected, though obviously the attenuation would be a controlling factor in practical applications. The writers, on the other hand, have given especial attention to this factor. Out of this research there emerged the remarkable fact that with hollow conducting guides there exists *one* and *only one* type of wave the attenuation of which *decreases* with increasing frequency; a unique characteristic which does not attach to dielectric wires, nor so far as the writers are aware, to any other type of guided wave.

IA. Transmission Through Hollow Conducting Cylinders

Throughout this paper it will be assumed that the cylindrical sheath possesses high conductivity and that the losses in the internal dielectric medium are either small or negligible. Subject to these assumptions the effect of dissipation on the attenuation of the wave is formulated in

[4] *Ann. der Phys.*, Vol. 49, 1916, pp. 907–933. This paper contains several collateral references.

[5] "Elektromagnetischen Wellen an Dielektrischen Drähten," *Ann. der Phys.*, Vol. 63, 1920, pp. 645–673.

[6] "Hyper-Frequency Wave Guides—General Considerations and Experimental Results," G. C. Southworth, this issue of the *Bell System Technical Journal*.

Section III. First, however, in the general discussion which immediately follows and, in particular, in the comparison with the usual guided wave transmission systems, attention will be confined to the ideal non-dissipative structure. This simplification brings out, in a simpler and more striking way, the peculiar transmission characteristics of the system, while, at the very high frequencies involved, it introduces negligible error except as regards the attenuation due to dissipation.

In the ordinary type of guided-wave systems, such for example, as that composed of two concentric conductors, or two parallel wires, the guiding conductors form two sides of a *circuit* in which equal and opposite currents flow, and the transverse lines of electric intensity terminate on the two sides of the circuit. In the system under consideration there is only *one* conductor and consequently there is no *circuit* in the usual sense. Corresponding to this difference in physical structure there are striking differences in the character of the waves propagated.

In the first place, in the ordinary type of guided wave system, the wave employed for the transmission of power and intelligence is the *Principal Plane Wave.* For the ideal non-dissipative case, the field of this wave is entirely transverse to the axis of the system; that is, the axial components of the electric and magnetic intensities are everywhere zero. Furthermore all frequencies are transmitted without attenuation with the same phase velocity; that of light in the medium. (Of course dissipation modifies the phenomena somewhat but in actual systems designed for efficient transmission the Principal Wave approximates to that just described.)

In the hollow conducting cylinder, on the other hand, *no principal transverse wave can exist;* that is, there must exist inside the cylinder either an axial component of the electric or the magnetic intensity, or both. Physically this is answerable to the absence of a circuit on which the transverse lines of force might terminate. Thus in the hollow conducting cylinder all the possible waves must be *complementary* waves; [7] a type which is ignored in the ordinary transmission system.

A second outstanding distinction is that in the hollow conducting cylinder, all frequencies below a critical frequency are attenuated while frequencies above the critical frequency are freely transmitted without attenuation.[8] In this respect the system behaves like a Campbell high-

[7] See "Guided and Radiated Energy in Wire Transmission," John R. Carson, *Jour. A.I.E.E.*, October 1924.

[8] It will be understood, of course, that this is strictly true only in the ideal case of no dissipation.

pass wave-filter. The exact value of the critical frequency depends, as shown later, on the type of wave transmitted; roughly speaking, however, the internal diameter must be approximately equal to one-half a wave-length in the internal dielectric medium at the lowest critical frequency. (The exact formula is diameter $> \frac{3.68}{2\pi}$ times the wave-length.) Since we are interested in freely transmitted waves it is evident at once that for a cylinder of practicable dimensions the frequencies employed must be relatively enormous. For this reason it may be appropriately said that the hollow conducting cylinder is applicable to the transmission of *hyper-frequency* waves alone.

The types of waves which can exist inside the cylinder are broadly classifiable as E-waves and H-waves.[9] By the term E-wave is to be understood a wave in which the axial component of the magnetic force is everywhere absent; correspondingly in the H-wave the axial component of the electric force is everywhere absent. In the E-waves the surface currents in the cylinder are entirely parallel to the axis thereof. On the other hand, in the H-waves the currents may have both transverse and axial components; that is, circulatory components around the periphery of the cylinder in planes normal to its axis as well as components parallel thereto.

In each class of wave there may exist a fundamental wave and in addition *geometrically* harmonic[10] waves. In the fundamental wave the phenomena do not vary around the periphery of the cylinder. In the nth harmonic wave (E_n- or H_n-wave) the phenomena vary around the periphery as $\cos n(\theta - \theta_n)$.

Each component E- or H-wave has its own individual critical frequency. Curiously enough the lowest critical frequency is possessed by the first harmonic H-wave; that is the H_1-wave. For this wave the critical frequency is given by the formula $d > \frac{3.68}{2\pi}\lambda$ where d is the internal diameter of the cylinder and λ the wave-length. In general, however, the critical frequency increases with the order of the harmonic.

In the usual transmission system, the transmission phenomena are determined and described in terms of the characteristic impedance and the propagation constant. By characteristic impedance the engineer understands the impedance actually presented by an infinitely long line to an electromotive force connected across the terminals of the circuit. Now since in the hollow conducting cylinder there is only one

[9] This terminology has been adopted as a matter of convenience. It is suggested by equations (1) where the field is expressed in terms of E_z and H_z. Another terminology is *transverse magnetic* and *transverse electric* waves.

[10] This term must not be confused with *frequency* harmonics.

conductor and hence no circuit, this concept breaks down. There is another way, however, in which the characteristic impedance may be defined, and by aid of which it remains a useful concept in hollow cylinder transmission. Writing $K = K_R + iK_I$ as the complex expression for the characteristic impedance, then it may be shown that

$$K = \bar{W} + i2\omega(\bar{T} - \bar{U}),$$

where $\bar{W}$ is the mean power transmitted, $\bar{T}$ is the mean stored magnetic energy, and $\bar{U}$ the mean stored electric energy, corresponding to an unit current. Now in the hollow conducting cylinder, for, say the E_0-wave, we can calculate

$$\bar{W} + i2\omega(\bar{T} - \bar{U})$$

for an *unit axial* current, and call this the characteristic impedance. Again for the H_0-wave we can calculate this quantity for an *unit circulating current per unit length* and designate it as the characteristic impedance. In addition, somewhat similar conventions apply to the harmonic waves.

One of the chief uses of the foregoing concept of characteristic impedance is in the calculation of the attenuation in the dissipative system. For, if corresponding to $\bar{W}$ we calculate the mean dissipation $\bar{Q}$ per unit length, then the attenuation α is given by

$$\alpha = \bar{Q}/2\bar{W}.$$

All actual systems are of course dissipative and consequently the wave is attenuated. If the hollow conducting cylinder were to be employed in practice for hyper-frequency wave transmission the securing of low and desirable attenuation characteristics would probably be the controlling consideration.

The attenuation in the free transmission range is due to (1) dissipation in the cylinder or sheath and (2) dissipation in the internal dielectric medium. The former is inherent and can be reduced only by employing a sheath of high conductivity and by properly designing the dimensions of the system. As regards the dielectric loss, this may be substantially eliminated by employing air as the dielectric medium. The use of a dielectric medium of high specific inductive capacity has the advantage of substantially reducing the critical frequency; on the other hand it inevitably introduces heavy losses and thus sharply increases the attenuation. The analysis of Section III brings out the remarkable fact that for the fundamental H-wave the attenuation decreases with increasing frequency; for all the other types it increases.

For the very high frequencies with which we shall be concerned in the following analysis, a physically very thin cylindrical metallic sheath behaves electrically as though it were infinitely thick. This fact greatly simplifies the mathematical treatment; its real importance, however, is that external interference is entirely eliminated.

As stated at the outset, this paper will not attempt to deal with the problem of the reflection phenomena which occur at the terminals of the system and at points of discontinuity. For a discussion of the general character of the boundary problem the reader is referred to "Guided and Radiated Energy in Wire Transmission."[7] It may be remarked here, however, that the simple engineering boundary conditions (continuity of current and potential) are entirely inadequate.

IB. Transmission Through Dielectric Guides

The greater part of this paper deals with transmission in thin hollow conducting cylinders; the last section, however, discusses briefly transmission along the dielectric wire.[3] Theoretically this type of transmission is extremely interesting and the mathematical theory resembles to a considerable extent that of hollow cylinder transmission. Unfortunately, however, dielectric losses are usually high. Hence our discussion of dielectric waves will be limited to a development of the fundamental equation from which the critical frequencies and the phase velocities can be determined.

II. Non-Dissipative Hollow Conducting Guides

In dealing with the propagation of hyper-frequency electromagnetic waves inside a long hollow conducting cylinder parallel to the z-axis, it is convenient to write the field equations in the appropriate cylindrical coordinates (ρ, θ, z) in the form,[11]

$$
\begin{aligned}
\lambda^2 H_\rho &= \frac{h^2}{\mu i \omega} \frac{1}{\rho} \frac{\partial}{\partial \theta} E_z - \gamma \frac{\partial}{\partial \rho} H_z, \\
\lambda^2 H_\theta &= -\frac{h^2}{\mu i \omega} \frac{\partial}{\partial \rho} E_z - \frac{\gamma}{\rho} \frac{\partial}{\partial \theta} H_z, \\
\lambda^2 E_\rho &= -\gamma \frac{\partial}{\partial \rho} E_z - \frac{\mu i \omega}{\rho} \frac{\partial}{\partial \theta} H_z, \qquad (1) \\
\lambda^2 E_\theta &= -\frac{\gamma}{\rho} \frac{\partial}{\partial \theta} E_z + \mu i \omega \frac{\partial}{\partial \rho} H_z, \\
\operatorname{div} E &= 0, \quad \operatorname{div} H = 0.
\end{aligned}
$$

[11] In this form the field is expressed explicitly in terms of the axial electric and magnetic intensities and their spatial derivatives. This is highly advantageous for the purposes of this paper.

In these equations the symbols have the following significance:

E_ρ, E_θ, E_z = components of electric force,
H_ρ, H_θ, H_z = components of magnetic force,
$\lambda^2 = \gamma^2 - h^2$,
γ = propagation constant,
$h^2 = \mu i\omega(4\pi\sigma + \epsilon i\omega/c^2) = 4\pi\sigma\mu i\omega - (\omega^2/v^2)$,
$v = c/\sqrt{\epsilon\mu}$ = velocity of light in the medium,
c = velocity of light in air,
μ = permeability of the medium in electromagnetic units,
σ = conductivity of the medium in electromagnetic units,
ϵ = dielectric constant of the medium in electrostatic units,
$\omega/2\pi$ = frequency,
$i = \sqrt{-1}$.

The solutions of these equations for the axial components of electric and magnetic force, E_z and H_z respectively, in the region, $0 \leqq \rho \leqq a$, a being the internal radius of the conductor, are of the form

$$\begin{aligned} E_z &= \sum_{n=0}^{\infty} J_n(\rho\lambda)(A_n \cos n\theta + B_n \sin n\theta) \exp.\,(i\omega t \pm \gamma z), \\ H_z &= \sum_{n=0}^{\infty} J_n(\rho\lambda)(C_n \cos n\theta + D_n \sin n\theta) \exp.\,(i\omega t \pm \gamma z), \end{aligned} \tag{2}$$

where A_n, B_n, C_n and D_n are arbitrary constants to be determined by boundary conditions and J_n is the Bessel function of the first kind or the internal Bessel function. The components of the transverse electromagnetic field may then be expressed by introducing (2) in (1).

We shall first discuss the simplest case, that in which there is no dissipation. The current will then be in a sheet on the surface, $\rho = a$, of the perfectly conducting cylinder. But the axial current density u_z and the circulating current density u_θ are given by

$$u_z = \frac{1}{4\pi} H_\theta, \quad \rho = a \tag{3}$$

and

$$u_\theta = \frac{1}{4\pi} H_z, \quad \rho = a. \tag{4}$$

Thus it follows that H_z and H_θ are discontinuous at the surface $\rho = a$ and the boundary conditions are simply $E_z = E_\theta = 0$. These conditions can be fulfilled by two types of waves: (1) a wave for which H_z

is zero everywhere, which will be called generically the E-wave and (2) a wave for which E_z is zero everywhere, which will be designated generically as the H-wave. (If the cylinder is dissipative, however, the E- and H-waves can exist alone only for the case of circular symmetry. In other words, unless $\partial/\partial\theta = 0$, neither the E_z nor the H_z component of the field can be identically zero. This will be discussed further in Section III.)

Assuming first a non-dissipative system, it will be seen that when H_z is zero everywhere,

$$E_z \text{ and } E_\theta \sim J_n(\lambda\rho) \begin{Bmatrix} \cos n\theta \\ \sin n\theta \end{Bmatrix}.$$

Thus the possible E-waves are determined by the boundary equation

$$J_n(\lambda a) = 0, \tag{5}$$

where

$$\lambda^2 = \gamma^2 + \omega^2/v^2.$$

This has an infinite number of real roots in λ determining an infinite number of possible waves. Only a finite number, m, of these waves will be unattenuated, however, for, if λ is to be real and γ pure imaginary, the frequency must be so high that

$$\omega/v > \lambda_{nm}, \tag{6}$$

where $\lambda_{nm}a$ is the mth root of $J_n(\lambda a) = 0$. It is therefore convenient to designate as the E_{nm}-wave that component of the E-wave for which

$$E_z \sim J_n(\lambda_{nm}\rho) \begin{Bmatrix} \cos n\theta \\ \sin n\theta \end{Bmatrix}.$$

Thus if

$$\lambda_{n,\, m+1} > \omega/v > \lambda_{nm},$$

the components $E_{n,\, m+1}$, $E_{n,\, m+2}$, $\cdots$ of the E-wave will all be attenuated but E_{n1}, E_{n2}, $\cdots$, E_{nm} will be unattenuated. There will also be only a finite number $n + 1$ of the components E_{01}, $E_{11}, \cdots E_{n1}$, for the frequency must be at least sufficiently high so that

$$\omega/v > \lambda_{n1},$$

where $\lambda_{n1}a$ is the lowest root (excluding zero) of $J_n(\lambda a) = 0$, in order to transmit the component E_{n1} of the E-wave without attenuation. Hence the E-wave consists of a doubly terminating series of possible components; for each of the finite number $k + 1$ possible values of n there will be m_n possible values of λa or a total of

$$m_0 + m_1 + m_2 + \cdots + m_k$$

possible modes of propagation.

For the H-wave, E_z is zero everywhere,

$$E_\theta \sim J_n'(\lambda\rho) \begin{Bmatrix} \cos n\theta \\ \sin n\theta \end{Bmatrix}$$

and the possible waves are determined by the transcendental equation

$$J_n'(\lambda a) = 0, \tag{7}$$

where

$$\lambda^2 = \gamma^2 + \omega^2/v^2$$

and $J_n'(z) = (d/dz)J_n(z)$. These values of λ and consequently of γ will, of course, differ from those characterizing the E-waves. Similarly, however, there will be a doubly terminating series of possible components, H_{nm}.

Hence for both types of wave the hollow conducting cylinder constitutes a high-pass wave-filter. The critical frequency f_{nm} of the E_{nm}-wave is given by

$$f_{nm} = r_{nm}(c/2\pi a\sqrt{\epsilon\mu}), \tag{8}$$

where r_{nm} is the mth root of $J_n(\lambda a) = 0$ or $r_{nm} = \lambda_{nm}a$. Similarly for the H_{nm}-wave, the critical frequency is

$$f_{nm}' = r_{nm}'(c/2\pi a\sqrt{\epsilon\mu}), \tag{8$'$}$$

where

$$r_{nm}' \text{ is the } m\text{th root of } J_{nm}'(\lambda a) = 0.$$

The propagation constant γ_{nm} is then

$$\gamma_{nm} = \frac{i\omega}{c} \cdot \frac{c}{v_{nm}'} = \frac{i\omega}{v_{nm}'}, \tag{9}$$

where the ratio c/v_{nm}' of the velocity of light in air to the phase velocity of the wave in response to any frequency f is given by

$$\begin{aligned} c/v_{nm}' &= \sqrt{\epsilon\mu}\sqrt{1 - (f_{nm}/f)^2} \\ &\to 0 \quad \text{when} \quad f \to f_{nm}, \\ &\to \sqrt{\epsilon\mu} \quad \text{when} \quad f \to \infty \end{aligned} \tag{10}$$

for the E-wave and

$$c/v_{nm}' = \sqrt{\epsilon\mu}\sqrt{1 - (f_{nm}'/f)^2}$$

for the H-wave.

For the E-wave we have

$$\begin{aligned} r_{01}, r_{02}, \cdots &= 2.405, 5.52, \cdots \\ r_{11}, r_{12}, \cdots &= 3.83, 7.02, \cdots \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

and for the H-wave

$$r_{01}', r_{02}', \cdots = 3.83, 7.02, \cdots$$
$$r_{11}', r_{12}', \cdots = 1.84, 5.33, \cdots.$$

Hence, it is possible to transmit a fundamental E-wave if the radius, dielectric constant, permeability and frequency are so related that,

$$fa\sqrt{\epsilon\mu} \geqq 2.405(c/2\pi), \tag{11}$$

a fundamental H-wave provided,

$$fa\sqrt{\epsilon\mu} \geqq 3.83(c/2\pi), \tag{12}$$

the component E_{11} of the E-wave provided

$$fa\sqrt{\epsilon\mu} \geqq 3.83(c/2\pi) \tag{13}$$

and the component H_{11} of the H-wave provided

$$fa\sqrt{\epsilon\mu} \geqq 1.84(c/2\pi). \tag{14}$$

Thus from the standpoint of minimum physical constants and dimensions the component H_{11} of the H-wave is most advantageous. The consideration of the attenuation characteristics below will show, however, that this advantage is outweighed, since in practice the attenuation will be the controlling factor.

We shall now consider the characteristic impedance of the system.[12] While the derivation of the characteristic impedance is interesting and valuable on its own merits, it also provides the basis for a quasi-synthetic and approximate method of deriving the attenuation which will be developed below. The results obtained here on the assumption of a perfect conductor will be valid in the dissipative case of the next section provided the conductivity is sufficiently high so that the relation, $4\pi\sigma \gg \epsilon\omega/c^2$, obtains among the constants of the sheath.

The characteristic impedance, K, is here defined as the transverse Complex Poynting Vector, P, integrated over the cross section of the system divided by the mean square current. Thus we have, in general,

$$\begin{aligned} P &= \frac{1}{8\pi}\int dS[E\cdot H^*], \\ &= \overline{W} + i2\omega(\overline{T} - \overline{U}), \end{aligned} \tag{15}$$

[12] See the discussion of the characteristic impedance in Section I of this paper. Equation (15) below is in agreement with the definition there given.

where $\overline{W}$ is the mean energy transmitted through the cylinder, $\overline{T}$ is the mean stored magnetic energy and $\overline{U}$ the mean stored electric energy, H^* denoting the conjugate imaginary of H. Then

$$K = K_R + iK_I \tag{16}$$

and

$$\tfrac{1}{2}K_R|\dot{I}|^2 = W, \tag{16a}$$

while

$$\tfrac{1}{2}K_I|I|^2 = 2\omega(\overline{T} - \overline{U}), \tag{16b}$$

I being the total current. (In a non-dissipative system $\overline{T} = \overline{U}$ and $K = K_R$.) Rewriting the integral in (15) we therefore have

$$W = \tfrac{1}{2}K_R|I|^2 = \frac{1}{8\pi}\left[\int_0^{2\pi}\int_0^a \rho(E_\rho H_\theta^* - E_\theta H_\rho^*)d\rho d\theta\right]_{\text{Real Part}}. \tag{17}$$

(From equations (1) it readily follows, that for any E- or H-wave, K may be made to depend upon either the transverse electric or transverse magnetic force alone by substituting in formula (17)

$$E_\rho H_\theta^* - E_\theta H_\rho^* = \frac{1}{c}\sqrt{\frac{\epsilon}{\mu}}\frac{v'}{v}[E]^2 = c\sqrt{\frac{\mu}{\epsilon}}\frac{v}{v'}[H]^2 \tag{18}$$

for the E-wave, and

$$E_\rho H_\theta^* - E_\theta H_\rho^* = \frac{1}{c}\sqrt{\frac{\epsilon}{\mu}}\frac{v}{v'}[E]^2 = c\sqrt{\frac{\mu}{\epsilon}}\frac{v'}{v}[H]^2 \tag{19}$$

for the H-wave, where $[E]^2$ and $[H]^2$ are defined as

$$[E]^2 = |E_\rho|^2 + |E_\theta|^2 \quad \text{and} \quad [H]^2 = |H_\rho|^2 + |H_\theta|^2.)$$

Consider first the fundamental E-wave. H_z, H_ρ and E_θ are zero and

$$\begin{aligned} E_z &= AJ_0(\rho\lambda), \\ E_\rho &= \frac{\gamma}{\lambda}AJ_1(\rho\lambda), \\ H_\theta &= \frac{\epsilon i\omega}{c^2}\frac{1}{\lambda}AJ_1(\rho\lambda), \end{aligned} \tag{20}$$

where

$$\lambda^2 = \gamma^2 + \omega^2/v^2$$

and

$$\lambda = r_{0m}/a. \quad (J_0(r_{0m}) = 0.)$$

From (3) the total axial current I_z in the sheath is given by

$$I_z = \frac{a}{2}H_\theta, \quad \rho = a. \tag{21}$$

Putting $I_z = 1$ then gives

$$A = \frac{c^2}{\epsilon i\omega} \frac{2\lambda}{aJ_1(\lambda a)} \cdot$$

Thus

$$\begin{aligned} K &= \frac{c}{\epsilon}\frac{c}{v'}\frac{2}{a^2}\frac{1}{(J_1(\lambda a))^2}\int_0^a \rho(J_1(\lambda\rho))^2 d\rho \\ &= \frac{c}{\epsilon}\cdot\frac{c}{v'}\left[1 + \left(\frac{J_0(\lambda a)}{J_1(\lambda a)}\right)^2 - \frac{2}{\lambda a}\frac{J_0(\lambda a)}{J_1(\lambda a)}\right] \\ &= c\sqrt{\mu/\epsilon}\,\sqrt{1-(f_{0m}/f)^2}. \end{aligned} \tag{22}$$

Now, for the fundamental component H_0 of the H-wave, E_z, E_ρ and H_θ are zero and

$$\begin{aligned} H_z &= CJ_0(\lambda\rho), \\ H_\rho &= \frac{\gamma}{\lambda} CJ_1(\lambda\rho), \\ E_\theta &= -\frac{\mu i\omega}{\lambda} CJ_1(\lambda\rho), \end{aligned} \tag{23}$$

where

$$\lambda^2 = \gamma^2 + \omega^2/v^2$$

and

$$\lambda = r_{0m}'/a. \qquad (J_0'(r_{0m}) = 0.)$$

There is no axial current transmitted by this wave but there is a circulating current in the sheath. From (4) this circulating current, I_θ, per unit length is given by

$$I_\theta = \frac{a}{2} H_z \quad \text{when} \quad \rho = a. \tag{24}$$

Thus, for the H_0-wave, we calculate the characteristic impedance with respect to unit circulating current per unit length of conductor. This gives

$$C = \frac{4\pi}{J_0(\lambda a)}$$

and

$$\begin{aligned} K &= \frac{1}{2v'}\left(\frac{4\pi\omega}{\lambda}\right)^2 \frac{1}{(J_0(\lambda a))^2}\int_0^a \rho(J_1(\lambda))^2 d\rho \\ &= (2\pi a)^2 \frac{\mu}{v'}\left(\frac{f}{r_{0m}'}\right)^2, \end{aligned} \tag{25}$$

where, as given above, r_{0m}' is the mth root of $J_0'(\lambda a)$, and, by (10),

$$v' = \frac{c}{\sqrt{\epsilon\mu}} \frac{1}{\sqrt{1-(f_{0m}'/f)^2}} \cdot$$

So we see that, while the characteristic impedance of the E_0-wave approaches a constant at very high frequencies, for the H_0-wave we have

$$K \sim \omega^2.$$

In other words, while the energy transmitted by the E_0-wave is independent of the frequency at sufficiently high values of frequency, that transmitted by the H_0-wave increases as the square of the frequency.

For the harmonic E- and H-waves, the currents vary as $\cos n\theta$ around the periphery of the sheath. Hence the total harmonic current is zero over any axial or normal cross-section. For these waves, however, it is possible and convenient to calculate the Complex Poynting Vector on the basis of the average mean square current intensities,

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\left|\frac{H_\theta}{4\pi}\right|^2 d\theta \quad \text{and} \quad \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\left|\frac{H_z}{4\pi}\right|^2 d\theta, \qquad \rho = a,$$

which we may assume for convenience to be of the same value, 1/2, as the mean square currents associated with the fundamental components.

On this basis we shall obtain first the characteristic impedance of any harmonic component E_n of the E-wave, ignoring dissipation. Putting

$$J_n(\lambda a) = 0 \quad \text{and} \quad \lambda a = r_{nm},$$

the Complex Poynting Vector becomes

$$\overline{W} = \frac{a^4}{16c}\sqrt{\frac{\epsilon}{\mu}}\frac{v'}{v}\left(\frac{\omega}{v'}\right)^2\frac{|A_n|^2 + |B_n|^2}{r_{nm}{}^2}(J_{n-1}(r_{nm}))^2. \tag{26}$$

On the basis of the current value which we are assuming

$$\frac{|A_n|^2 + |B_n|^2}{\lambda^2}(J_{n-1}(r_{nm}))^2 = 32\pi^2\left(\frac{c^2}{\epsilon\omega}\right)^2. \tag{27}$$

Thus

$$K_R = (2\pi a)^2 c\sqrt{\mu/\epsilon}\sqrt{1 - (f_{nm}/f)^2}. \tag{28}$$

Similarly, for the component H_n of the H-wave, we put

$$J_n'(\lambda a) = 0 \quad \text{and} \quad \lambda a = r_{nm}',$$

getting

$$\overline{W} = \frac{a^4 c}{16}\sqrt{\frac{\mu}{\epsilon}}\frac{v'}{v}\left(\frac{\omega}{v'}\right)^2(|C_n|^2 + |D_n|^2)\frac{(J_n(r_{nm}'))^2}{(r_{nm}')^2}\left(1 - \frac{n^2}{(r_{nm}')^2}\right). \tag{29}$$

where

$$(|C_n|^2 + |D_n|^2)(J_n(r_{nm}'))^2 = 32\pi^2. \tag{30}$$

Thus

$$K_R = (2\pi a)^4 \left(1 - \frac{n^2}{(r_{nm}')^2}\right)\frac{\mu}{v'}\left(\frac{f}{r_{nm}'}\right)^2, \tag{31}$$

where, as given above, r_{nm}' is the mth root of $J_n'(\lambda a)$ and, by (10)

$$v' = \frac{c}{\sqrt{\epsilon\mu}}\frac{1}{\sqrt{1 - (f_{nm}'/f)^2}}.$$

Thus the mean transmitted energy and the characteristic impedance of all components of the H-wave increase as the square of the frequency whereas these characteristics of the E-wave are constant with respect to frequency. To appreciate the bearing of this difference upon the comparative attenuations consider the following argument.

Since the wave varies along the z-axis of the transmission system as exp. $((-\alpha - i\beta)z)$, α and β denoting the attenuation and phase constants per unit length, respectively,

$$\frac{\partial \overline{W}}{\partial z} = -2\alpha\overline{W}. \tag{32}$$

But, denoting by Q the dissipation loss per unit length of the transmission system, we have also

$$\frac{\partial \overline{W}}{\partial z} = -\overline{Q}. \tag{33}$$

Hence,

$$\alpha = \overline{Q}/2\overline{W} \tag{34}$$

$$= (4\pi\overline{Q}\textstyle\int dS[E\cdot H^*]_z)_{\text{Real Part}}. \tag{35}$$

Thus, we see that, if the mean dissipation loss, $\overline{Q}$, is known or readily obtainable, the Complex Poynting Vector, $\overline{W}$, leads immediately to the attenuation.

To obtain $\overline{Q}$ we have the formula

$$\overline{Q} = -\left(\frac{1}{8\pi}\int dS[E\cdot H^*]_r\right)_{\text{Real Part}}. \tag{36}$$

Thus α may also be written

$$\alpha = \left(\frac{-\int dS[E\cdot H^*]_r}{2\int dS[E\cdot H^*]_z}\right)_{\text{Real Part}} \tag{37}$$

in which it is evident the current is not explicitly involved. If we write

$$\bar{Q} = R(I^2)_m$$

and

$$\bar{W} = K_R(I^2)_m,$$

R being the resistance per unit length and K_R the characteristic impedance with respect to the mean square current $(I^2)_m$ we have in addition

$$\alpha = R/2K_R. \tag{38}$$

Before continuing our discussion of attenuation, we shall, therefore, have to calculate the losses in the sheath and the internal dielectric medium.

III. Dissipative Hollow Conducting Guides

In the ideal case of the preceding section, where the conductivities σ_1 and σ_2 of the dielectric and conductor are, respectively, zero and infinity, the boundary conditions are simply that $E_z = E_\theta = 0$ at the surface, $\rho = a$. When we take into account the dissipation which is actually present in the conductor (and the dielectric as well) the boundary conditions are the continuity of both the tangential electric and tangential magnetic forces. This double set of boundary conditions makes the problem inherently more difficult, of course. As we are assuming a good conductor and dielectric, we shall treat the dissipative case as a departure of the first order from the ideal case. Thus, since the dissipation has a negligible first order effect upon the phase velocity, the propagation constant γ will now be

$$\gamma = i\omega/v' + \alpha,$$

where α denotes the attenuation.

We must now consider the field in the sheath as well as the field in the inner dielectric medium. When necessary we distinguish between the electrical constants of the two media by the subscripts 2 and 1, respectively. We suppose that the sheath is electrically very thick, a legitimate assumption at the very high frequencies in which we are interested, and write for $\rho > a$,

$$\begin{aligned} E_z &= \sum_{n=0}^{\infty} K_n(\rho\lambda_2)(A_n' \cos n\theta + B_n' \sin n\theta) \exp. (i\omega t \pm \gamma z), \\ H_z &= \sum_{n=0}^{\infty} K_n(\rho\lambda_2)(C_n' \cos n\theta + D_n' \sin n\theta) \exp. (i\omega t \pm \gamma z), \end{aligned} \tag{39}$$

where

$$\lambda_2^2 = \gamma^2 - h_2^2$$

and K_n is the Bessel function of the second kind [13] (or the external Bessel function) and obtain H_ρ, H_θ, E_ρ and E_θ from (39) and (1). Putting

$$\lambda_1 a = y \quad \text{and} \quad \lambda_2 a = x,$$

and equating the tangential electric and magnetic forces E_z, E_θ and H_z, H_θ at the boundary surface $\rho = a$, we obtain eight homogeneous equations in the eight arbitrary constants. A non-trivial solution requires the vanishing of the determinant; this condition leads to the transcendental equation:

$$\left(\frac{h_1^2}{\mu_1}\frac{J_n'(y)}{yJ_n(y)} - \frac{h_2^2}{\mu_2}\frac{K_n'(x)}{xK_n(x)}\right)\left(\mu_1\frac{J_n'(y)}{yJ_n(y)} - \mu_2\frac{K_n'(x)}{xK_n(x)}\right) - n^2\gamma^2\left(\frac{1}{y^2} - \frac{1}{x^2}\right)^2 = 0, \quad (40)$$

where

$$y^2 = a^2(\gamma^2 - h_1^2) \quad (40a)$$

and

$$x^2 = a^2(\gamma^2 - h_2^2). \quad (40b)$$

The propagation constant γ is then determined by equation (40).

We mentioned in Section II that the E- or H-waves cannot exist alone in the dissipative case unless they are circularly symmetrical and it may be noticed that both E_z and H_z were required in the analysis of the preceding paragraph. To show that E_z and H_z must coexist when the conductor is dissipative, assume for the moment that $E_z = 0$. The boundary equations when $n \neq 0$ are then

$$\begin{aligned} C_nJ_n(y) &= C_n'K_n(x), & D_nJ_n(y) &= D_n'K_n(x), \\ \frac{C_n}{y^2}J_n(y) &= \frac{C_n'}{x^2}K_n(x), & \frac{D_n}{y^2}J_n(y) &= \frac{D_n'}{x^2}K_n(x), \\ \frac{\mu_1 C_n}{y}J_n'(y) &= \frac{\mu_2 C_n'}{x}K_n'(x), & \frac{\mu_1 D_n}{y}J_n'(y) &= \frac{\mu_2 D_n'}{x}K_n'(x); \end{aligned} \quad (41)$$

six equations which cannot be satisfied by four arbitrary constants. When $n = 0$, however, H_θ is everywhere zero and the boundary equations are simply

$$\begin{aligned} C_0J_0(y) &= C_0'K_0(x), \\ \frac{\mu_1 C_0}{y}J_0'(y) &= \frac{\mu_2 C_0'}{x}K_0'(x). \end{aligned} \quad (42)$$

[13] This is the Hankel function given in Jahnke und Emde, "Funktionentafeln," p. 94, 1st ed., and denoted by $H_n^{(1)}(z)$ when $\arg z < \pi$. To avoid confusion with the nth harmonic of the H-wave, we shall use K_n as a generic symbol to denote the external Bessel function.

Similarly the boundary equations can be satisfied when $H_z = 0$ provided $n = 0$ but not when $n \neq 0$.

Although E_z and H_z must co-exist in the dissipative case, one or the other will predominate in the actual wave provided the conductivity is so high that $4\pi\sigma_2 \gg \epsilon_2\omega/c^2$, a condition which is true of a good conductor unless $f \to \infty$. That this is so or, in other words, that the actual wave approximates either an E- or an H-wave will now be shown from equation (40). Since it is assumed that the conductivity is high or that

$$4\pi\sigma_2 \gg \epsilon_2\omega/c^2 \quad \text{and} \quad h_2^2 \gg \gamma^2, \tag{43}$$

$x = a\sqrt{-4\pi\sigma_2\mu_2 i\omega}$ and the asymptotic values of $K_n(x)$ and $K_n'(x)$ are valid. Equation (40) may then be written

$$\left(\frac{h_1^2}{\mu_1}\frac{J_n'(y)}{yJ_n(y)} - \frac{h_2}{\mu_2 a}\right)\left(\mu_1\frac{J_n'(y)}{yJ_n(y)} - \frac{\mu_2}{ah_2}\right) - n^2\gamma^2\left(\frac{1}{y^2} + \frac{1}{a^2h_2^2}\right) = 0. \tag{44}$$

When $h_2 = \infty$, (44) reduces to

$$J_n'(y) = 0 \text{ provided } J_n(y) \neq 0 \tag{45}$$

and to

$$J_n(y) = 0 \text{ provided } J_n'(y) \neq 0. \tag{46}$$

Thus there are two possible solutions of (44). These are in the neighborhood of $y = r$ and of $y = r'$, where r and r', respectively, are roots of $J_n(y) = 0$ and of $J_n'(y) = 0$, the equations characterizing the E- and the H-wave, respectively. We shall, therefore, refer to E- and H-waves in the dissipative case with the understanding that the actual wave approximates either one or the other type in a cylinder of sufficiently high conductivity.

As stated above, the propagation constant γ may be determined by solving equation (40). The procedure is straightforward but is complicated by the necessity for approximations and does not easily admit of physical interpretation. We may obtain the same attenuation formulas by means of the quasi-synthetic method developed at the end of Section II.

The high-frequency attenuation of the symmetric E- and H-waves is easily derived from equation (38). Here R, the resistance per unit length of the cylinder for the E-wave at sufficiently high frequencies, is given by

$$R = \frac{\sqrt{\mu_2 f/\sigma_2}}{a}. \tag{47}$$

Introducing K of (22) for K_R, and understanding that $\epsilon = \epsilon_1$, while it is assumed that $\epsilon_2 = 1$, we have

$$\alpha = \frac{1}{2ac}\sqrt{\frac{\epsilon\mu_2 f}{\mu_1\sigma_2}}\frac{1}{\sqrt{1-(f_{0m}/f)^2}} \tag{48}$$

to a high precision at high frequencies ($f > f_{0m}$). This is, of course, the contribution of the conductor and ignores the effect of the conductivity of the dielectric.

Similarly, for the fundamental H-wave, the resistance per unit length of the cylinder at sufficiently high frequencies, from equations (1) and (39) and the relations

$$\overline{Q} = \left[\frac{a}{8\pi}\int_0^{2\pi} E_\theta H_z^* d\theta\right]_{\text{Real Part}}$$

and

$$(I^2)_m = \frac{1}{2}\,|I_\theta|^2 = \frac{1}{2},$$

is given by

$$R = \frac{\sqrt{\mu_2 f/\sigma_2}}{a}. \tag{49}$$

Putting K of (25) for K_R, gives, to the same precision as (48), when $f > f_{0m}'$,

$$\alpha = \frac{\sqrt{\epsilon\mu_2/\mu_1\sigma_2}(f_{0m}')^2}{2ac}\frac{f^{-3/2}}{\sqrt{1-(f_{0m}'/f)^2}}. \tag{50}$$

Formulas (48) and (50), respectively, may be written in the form

$$\alpha = \frac{\alpha_0}{\sqrt{1-(f_c/f)^2}}, \qquad f > f_c \tag{48$'$}$$

and

$$\alpha = \alpha_0\frac{(f_c'/f)^2}{\sqrt{1-(f_c'/f)^2}}, \qquad f > f_c' \tag{50$'$}$$

where

$$\alpha_0 = \frac{1}{2ac}\sqrt{\frac{\epsilon\mu_2 f}{\mu_1\sigma_2}}$$

and f_c and f_c' are the critical frequencies of the fundamental E- and H-waves, respectively, as given by (8) and (8)$'$.

Thus, in the neighborhood of their respective critical frequencies, the attenuations of the two waves are functionally the same; ultimately, however, while the attenuation of the fundamental E-wave *increases* as $f^{1/2}$, the attenuation of the fundamental H-wave *decreases* as $f^{-3/2}$; a remarkable property peculiar to this type of wave alone.

By extending the preceding treatment to the harmonic waves, it is found after some rather laborious analysis that for all the component E-waves,

$$\alpha = \frac{\alpha_0}{\sqrt{1 - (f_n/f)^2}}, \qquad f > f_n. \tag{51}$$

Care must be taken, of course, to choose the correct critical frequency ($f_n = f_{nm}$) for the particular component wave under consideration.

For all the H-waves (including the fundamental H-wave) it is found that

$$\alpha = \frac{\alpha_0}{\sqrt{1 - (f_n'/f)^2}} \left((f_n'/f)^2 + \frac{(n/r')^2}{1 - (n/r')^2} \right). \tag{52}$$

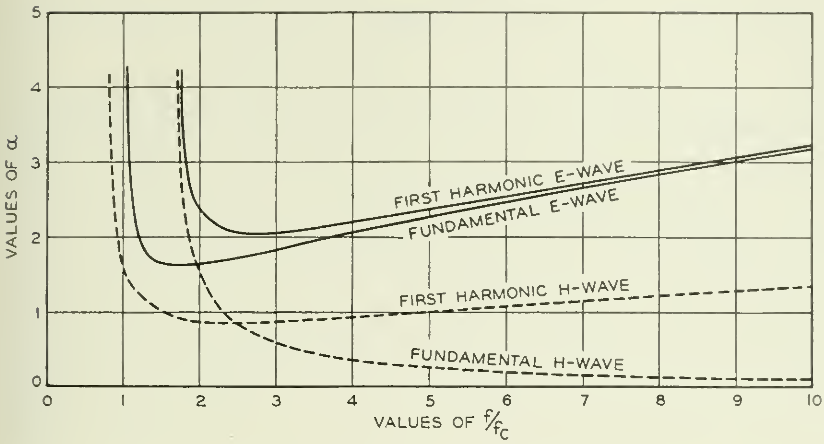
Here n is the order of the geometric harmonic wave (H_n-wave) and r' is the root of $J_n'(y)$ corresponding to the particular component wave under consideration.

The foregoing formulates the attenuation due to dissipation in the sheath alone. If we suppose that the dielectric has a very small but finite conductivity σ_1, then there must be added to the attenuation, for all types of waves, a term

$$\frac{2\pi\sigma_1 c\sqrt{\mu_1/\epsilon}}{\sqrt{1 - (f_n/f)^2}}. \tag{53}$$

To a first order approximation the dissipation has no effect on the phase velocity, which is simply v'.

Comparative values of attenuation are shown on the accompanying drawing for the fundamental and for the first harmonic E- and H-waves. This is the attenuation due to the loss in the conductor only. That due to the dielectric loss, the term given by (53), must be added. In many instances, we cannot say how large this term will be, for the losses in many dielectrics at the high frequencies involved herein are not known with any certainty at present. Such approximate calculations as we have made, however, have shown them to be very large except in the case of air.



Attenuation, α, in Hollow Conducting Cylinder.

Multiply ordinates by $A_0 = \dfrac{4.66}{d}\sqrt{\dfrac{f_c}{\sigma \times 10^4}}$ to read db per mile.

For copper, $A_0 = \dfrac{1.89\sqrt{f_c}}{d}$.

Multiply abscissae by $f_c = (2.30/d)10^4$ to read frequency in megacycles. Here
f_c = critical frequency of fundamental E-wave in megacycles,
d = inner diameter of cylinder in centimeters,
σ = conductivity of cylinder in emu
= 6.06×10^{-4} for copper.

IV. Dielectric Cylindrical Guides

We shall now pass to the mathematical theory of waves in dielectric "wires" of circular cross-section, immersed in air. We assume that the dielectric is perfect. The field in such a dielectric guide, and in the air outside, can be represented by the same general expressions as in hollow tubes. Thus for the nth harmonic wave, we have

$$\begin{aligned} E_z &= A_n J_n(\lambda_1\rho)\cos n\theta, \qquad H_z = B_n J_n(\lambda_1\rho)\sin n\theta, \text{ in the guide,} \\ E_z &= C_n K_n(\lambda_2\rho)\cos n\theta, \qquad H_z = D_n K_n(\lambda_2\rho)\sin n\theta, \text{ in the air.} \end{aligned} \tag{54}$$

The exponential factor $e^{-\gamma z + i\omega t}$ is implied in these as well as in the subsequent expressions for the field intensities. Another fundamental solution is obtained by changing θ into $\theta + \pi/2n$.

The transverse components of E and H are obtainable from E_z and H_z by differentiation. For our present purposes we need only E_θ and H_θ; these are

$$
\begin{aligned}
E_\theta &= \left[A_n \frac{n\gamma}{\lambda_1^2 \rho} J_n(\lambda_1 \rho) + B_n \frac{i\omega\mu_1}{\lambda_1} J_n'(\lambda_1 \rho) \right] \sin n\theta, \text{ in the guide,} \\
H_\theta &= - \left[A_n \frac{i\omega\epsilon_1}{\lambda_1 c^2} J_n'(\lambda_1 \rho) + B_n \frac{n\gamma}{\lambda_1^2 \rho} J_n(\lambda_1 \rho) \right] \cos n\theta, \text{ in the guide,} \\
E_\theta &= \left[C_n \frac{n\gamma}{\lambda_2^2 \rho} K_n(\lambda_2 \rho) + D_n \frac{i\omega\mu_2}{\lambda_2} K_n'(\lambda_2 \rho) \right] \sin n\theta, \text{ in the air,} \\
H_\theta &= - \left[C_n \frac{i\omega\epsilon_2}{\lambda_2 c^2} K_n'(\lambda_2 \rho) + D_n \frac{n\gamma}{\lambda_2^2 \rho} K_n(\lambda_2 \rho) \right] \cos n\theta, \text{ in the air.}
\end{aligned}
\tag{55}
$$

The boundary conditions require the continuity of the tangential components of E and H. Hence if a is the radius of the guide, we have

$$
A_n J_n(\lambda_1 a) = C_n K_n(\lambda_2 a), \quad B_n J_n(\lambda_1 a) = D_n K_n(\lambda_2 a), \tag{56}
$$

$$
A_n \frac{n\gamma}{\lambda_1^2 a} J_n(\lambda_1 a) + B_n \frac{i\omega\mu_1}{\lambda_1} J_n'(\lambda_1 a) = C_n \frac{n\gamma}{\lambda_2^2 a} K_n(\lambda_2 a) + D_n \frac{i\omega\mu_2}{\lambda_2} K_n'(\lambda_2 a),
$$

$$
A_n \frac{i\omega\epsilon_1}{\lambda_1 c^2} J_n'(\lambda_1 a) + B_n \frac{n\gamma}{\lambda_1^2 a} J_n(\lambda_1 a) = C_n \frac{i\omega\epsilon_2}{\lambda_2 c^2} K_n'(\lambda_2 a) + D_n \frac{n\gamma}{\lambda_2^2 a} K_n(\lambda_2 a).
$$

This is a homogeneous set of linear equations in the coefficients A, B, C and D from which only the ratios of these coefficients can be determined. But there are only *three* independent ratios and *four* equations; eliminating these ratios we shall obtain the *characteristic equation* of our boundary value problem from which the propagation constant γ can be calculated in terms of the frequency, the radius of the guide and the electromagnetic constants of the guide.

If $n = 0$, the above set of equations breaks up into two independent sets connecting the pairs A, C and B, D. Hence non-trivial solutions are possible by letting $A = C = 0$ or $B = D = 0$. In one case E_z is zero everywhere and in the other H_z vanishes. Thus in the circularly symmetric case we have waves of either the E-type or H-type in the sense previously defined. But if $n \neq 0$, then E_z and H_z must be present simultaneously.

The case $n = 0$ is so much simpler than the others that we shall examine it separately. Thus the characteristic equation for an E_0-wave is

$$
\frac{\epsilon_1 J_1(\lambda_1 a)}{\lambda_1 a J_0(\lambda_1 a)} = \frac{\epsilon_2 K_1(\lambda_2 a)}{\lambda_2 a K_0(\lambda_2 a)}, \tag{57}
$$

and that for an H_0-wave is

$$
\frac{\mu_1 J_1(\lambda_1 a)}{\lambda_1 a J_0(\lambda_1 a)} = \frac{\mu_2 K_1(\lambda_2 a)}{\lambda_2 a K_0(\lambda_2 a)}. \tag{58}
$$

In addition to either of these equations, we have

$$\gamma = \sqrt{\lambda_1^2 - \mu_1\epsilon_1\omega^2/c^2} = \sqrt{\lambda_2^2 - \mu_2\epsilon_2\omega^2/c^2}, \tag{59}$$

and the condition that for truly guided waves γ and λ_2 must be pure imaginary while λ_1 is real. When λ_2 is pure imaginary, the Hankel function of the second kind will decrease almost exponentially with increasing distance from the guide if this distance is sufficiently large.

If λ_1 and λ_2 are taken from (59) and substituted in (57) and (58) we shall have equations determining γ in terms of ω. Unfortunately these equations do not admit of an explicit solution for γ. It is possible, however, to carry out the numerical calculations in the following manner. We plot the left and the right terms of (57), let us say, against their arguments; then we select a pair of values of these arguments corresponding to equal ordinates. Let us suppose that we obtain

$$(\lambda_1 a)^2 = p^2, \quad (\lambda_2 a)^2 = -q^2, \tag{60}$$

where p and q are real. Referring to section III, we have $p = y$ and $iq = x$. Substituting these in (59) and solving, we have

$$\omega = \frac{c\sqrt{p^2 + q^2}}{a\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}}, \quad \gamma = i\sqrt{\frac{\mu_2\epsilon_2\omega^2}{c^2} + \frac{q^2}{a^2}}. \tag{61}$$

Since μ_1 usually equals μ_2, the guided waves are possible only if the dielectric constant of the guide is higher than that of the surrounding medium.

The lowest value of q is zero; the right member of (57) is then infinite and the corresponding value of p must then be a root of

$$J_0(p_m) = 0. \tag{62}$$

Corresponding to each root we have a different mode of propagation. The lowest frequency which can be transmitted in any particular mode and the corresponding propagation constant are given by

$$\omega_m = \frac{cp_m}{a\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}}, \quad \gamma = \frac{i\omega\sqrt{\mu_2\epsilon_2}}{c}. \tag{63}$$

At this frequency the phase velocity of propagation is equal to that of light in air. Since λ_2 is small, the field extends to great distances outside the guide. As q increases indefinitely, the right part of (57) approaches zero and p must approach the root of $J_1(x)$ near the particular root of J_0 that we happen to be considering. Thus for large values of q, we have approximately

$$\omega = \frac{cq}{a\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}}, \quad \gamma = \frac{i\omega\sqrt{\mu_1\epsilon_1}}{c}. \tag{64}$$

Hence at high frequencies the propagation takes place substantially with the velocity of light appropriate to the substance of the guide. The constant λ_2 being large, the field is concentrated largely in the guide.

Returning to the general n-th harmonic wave, we set

$$\begin{aligned} A_n &= SK_n(\lambda_2 a), \quad C_n = SJ_n(\lambda_1 a), \\ B_n &= TK_n(\lambda_2 a), \quad D_n = TJ_n(\lambda_1 a). \end{aligned} \tag{65}$$

Substitute in the last two equations of (56) and eliminate S and T. Thus we obtain

$$\begin{aligned} \frac{n\gamma}{a} J_n K_n \left(\frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}\right) S &= i\omega \left(\frac{\mu_2 J_n K_n'}{\lambda_2} - \frac{\mu_1 K_n J_n'}{\lambda_1}\right) T, \\ \frac{i\omega}{c^2}\left(\frac{\epsilon_2 J_n K_n'}{\lambda_2} - \frac{\epsilon_1 K_n J_n'}{\lambda_1}\right) S &= \frac{n\gamma}{a} J_n K_n \left(\frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}\right) T. \end{aligned} \tag{66}$$

Subsequently

$$\begin{aligned} \frac{\epsilon_1\mu_1 J_n'^2}{p^2 J_n^2} - \frac{i(\epsilon_1\mu_2 + \mu_1\epsilon_2) J_n' K_n'}{pq J_n K_n} - \frac{\epsilon_2\mu_2 K_n'^2}{q^2 K_n^2} \\ = n^2 \left(\frac{1}{p^2} + \frac{1}{q^2}\right)\left(\frac{\epsilon_1\mu_1}{p^2} + \frac{\epsilon_2\mu_2}{q^2}\right), \end{aligned} \tag{67}$$

and finally

$$\begin{aligned} \epsilon_2\mu_2 \frac{K_{n-1}K_{n+1}}{q^2 K_n^2} - \epsilon_1\mu_1 \frac{J_{n-1}J_{n+1}}{p^2 J_n^2} - \frac{i(\epsilon_1\mu_2 + \mu_1\epsilon_2) J_n' K_n'}{pq J_n K_n} \\ = n^2 \frac{\epsilon_1\mu_1 + \epsilon_2\mu_2}{p^2 q^2}. \end{aligned} \tag{68}$$

Allowing q to approach zero, we shall obtain in the limit an equation whose roots in conjunction with (61) determine the critical frequencies.

Thus if $n > 1$, we obtain

$$(\epsilon_1\mu_2 + \mu_1\epsilon_2)\frac{pJ_{n-1}(p)}{J_n(p)} = n(\epsilon_1 - \epsilon_2)(\mu_2 - \mu_1) + \frac{\epsilon_2\mu_2}{n-1} p^2. \tag{69}$$

Since ordinarily $\mu_1 = \mu_2$, (69) becomes

$$\frac{J_{n-1}(p)}{pJ_n(p)} = \frac{\epsilon_2}{(n-1)(\epsilon_1 + \epsilon_2)}. \tag{70}$$

If the dielectric constant of the guide is very much higher than that of the surrounding air, the first few roots of (70) are very close to those of $J_{n-1}(p) = 0$. As q increases indefinitely (68) degenerates into

$$\frac{J_{n-1}(p)J_{n+1}(p)}{J_n^2(p)} = 0. \tag{71}$$

Thus in the limit the roots of (68) will be *exactly* those of $J_{n-1}(p) = 0$. In other words as q varies from 0 to ∞ the corresponding value of p as given by (68) will not change much. It might appear that the limiting values of p could be roots of $J_{n+1}(p) = 0$; this is not possible, however, because in the process of transition p would have to pass through the intermediate zero of $J_n(p)$ and no real value of q is consistent with such zeros.

The case $n = 1$ requires a special examination. After multiplying (68) by q^2 and permitting q to approach zero, we find that the first term tends to infinity while the last term becomes a constant. Since the limit of $\frac{qK_1'}{K_1}$ is finite, $J_1(p)$ must approach zero. Thus for $n = 1$, the critical frequencies are determined by the zeros of $J_1(p)$.

One interesting point may be mentioned in conclusion. If the guide were surrounded by a hypothetical medium of zero dielectric constant, equation (57) for the E_0-waves would become

$$\frac{J_1(\lambda_1 a)}{\lambda_1 a J_0(\lambda_1 a)} = 0, \quad J_1(\lambda_1 a) = 0. \tag{72}$$

Thus the critical frequencies would be given by the roots of $J_1(p) = 0$ and not by those of $J_0(p) = 0$ as is the case for *any* ratio $\frac{\epsilon_2}{\epsilon_1}$ *different from zero* no matter how small it may be. Our first impression is that this result does violence to our physical common sense which demands that the hypothetical idealized case should be an approximation to the real one when one dielectric constant is large in comparison with the other. And indeed common sense is justified if one does not adhere too closely to the exact mathematical definition of the expression "critical frequency." In the region between any particular zero of $J_0(p)$, giving the true critical frequency, and the corresponding zero [14] of $J_1(p)$, giving the "approximate" critical frequency, most of the energy travels *outside* the guide, with a velocity substantially equal to that of light in the surrounding medium. The "approximate" critical frequency marks the region of the most rapid transition from wave propagation outside the guide to that inside the guide.

[14] This zero is always larger than that of $J_0(p)$.

A Magneto-Elastic Source of Noise in Steel Telephone Wires

By W. O. PENNELL and H. P. LAWTHER

The appearance of an electromotive force at the terminals of a vibrating rod or wire of magnetic material was investigated. It was concluded from experiments somewhat more simple and direct than those employed by other investigators that the effect was due to changes in the state of circular magnetization accompanying the variations of stress. The results suggested problems for more intensive investigation and applications of possible practical value.

THERE probably are few persons who have not had the experience of standing near some telephone or telegraph line out in the open and hearing the singing of the wires resulting from the wind blowing over them. Perhaps in childhood it was a source of wonderment why these sounds could not be heard at the telephones, and later, upon learning that telephone transmission was accomplished electrically and that these vibrations were mechanical only, tolerant amusement was felt at this earlier ingenuousness. Apparently it has remained until a very recent date for the discovery unmistakably to be made that it is possible under certain conditions for the mechanical vibration of a telephone wire to generate electromotive forces of sufficient magnitude to become objectionably audible in the telephone circuit. It was in April, 1935 that Mr. G. G. Jones of the Long Lines Department of the American Telephone and Telegraph Company mentioned to one of the writers the experience his Company had had a short time before in tracing down a supposed case of inductive noise to the action of the wind upon a 1200-foot steel-wire river-crossing span near Topeka, Kansas. This particular case had been cleared promptly by the application of suitable vibration damping devices generally recommended for situations where vibration might cause trouble. Special investigations then were made of long steel-wire spans at several locations in the Southwestern Bell Telephone Company territory, and it was revealed that some slight noise derived from this source actually was present in every case—and in one particular instance, where the wind velocity and direction happened to become very favorable to the production of wire vibration during the time of the inspection, the noise arose to a serious magnitude. In none of these locations had there been previous evidence that vibration was serious. That so simple and direct a phenomenon had escaped identification at the hands of telephone workers through the years of the art's existence

seemed remarkable, and especially so in view of the fact that Bell, the inventor of the telephone, in 1879 made passing note [1] of an experimental finding that probably was due in part to this effect. Accordingly, the writers' interest was aroused to the extent that an investigation was undertaken to learn the basic cause of the observed result.

In the light of subsequent knowledge it was surprising that some keen observer had not predicted and demonstrated the effect as the natural and necessary consequence of the researches of Ewing [2] and his predecessors upon the relation between state of magnetization and state of stress of a ferromagnetic specimen. Apparently it remained for von Hippel and Stierstadt [3] first to remark the phenomenon in 1931. These men were unable to interpret the effect in simple terms, however, and their reports presented a series of premature conclusions. Von Auwers [4] alone recognized the effect as capable of complete and satisfying explanation on the basis of magneto-elasticity, but he chose a method of establishing this, the interpretation of which was quite involved. For their own satisfaction in comprehending the phenomenon the present writers were led to conduct a series of experiments of qualitative character. It was felt that knowledge and clear understanding of the effect should be of immediate interest and value to workers in the general field of electrical communication in the United States.

With the aid of an amplifier having a gain of 110 decibels and terminating in a loud speaker it was practicable to conduct the experimentation with specimens of table-top dimensions. This amplifier had an input impedance very much higher than that of any of the specimens, and the response of the speaker therefore was proportional to the voltages generated by the specimens. A stretched iron wire three feet long would yield a clear sound in the speaker when its ends were connected to the input of the amplifier while it was being mechanically stimulated by plucking or bowing, and the sound from the speaker would be closely of the same quality as that heard by direct air transmission from the vibrating specimen. With this arrangement it was possible to detect any change in the magnitude of the effect as great as two to one simply by observing the loudness of the sound.

It was verified immediately that the effect must be dependent upon the property of ferro-magnetism. Taut wires of soft iron, tempered steel, or nichrome; rods of soft steel or permalloy—all produced strong sounds in the speaker when their ends were connected to the input of the amplifier while they were stimulated to vibration by bowing, plucking, or tapping. With wires of copper or brass, or with a rod of

carbon, no sound could be heard. Those small electromotive forces which must have resulted from the motion of any of the specimens in the earth's magnetic field were totally inappreciable with the apparatus employed.

Now the appearance of a potential difference between the two ends of a wire consequent to its vibration necessarily must have implied one of two situations—either there was some external influence or some relation to its surroundings which was capable of discriminating between the two ends, or else the wire possessed inherently some property that differentiated between the two directions along its length. Accordingly, exhaustive efforts were made to learn if orientation had any influence on the phenomenon. A stretched soft iron wire about three feet long yielding a clear sound in the speaker upon being plucked was employed. First, the ends of the wire were held stationary, and the wire was plucked time and again so that its plane of vibration covered representatively the various inclinations possible for this. Then there was tried a large number of positions for the axis of the wire, spread uniformly over the complete sphere. No response to orientation could be found. This negative result meant that the phenomenon under investigation must have arisen fundamentally through some condition of polarity resident within the specimen.[5]

With the phenomenon associated so definitely with the ferromagnetic property of the substance, and attributable so certainly to some quality of polarity of the specimen, it was but natural to recall [2] the considerable changes in their states of magnetization which accompany the applications of stresses with ferro-magnetic materials. In order to have produced an electromotive force along the axis, the state of magnetization of a rod or wire would have had to change in that same sense in which magnetization would have been acquired when an electromotive force was applied, and current allowed to flow, between its two ends; i.e., magnetization in closed circular paths centered upon the axis and at right angles thereto. Having formed this reasonable hypothesis of the fundamental process, experiment then was carried along the lines of testing its validity.

Any circular magnetization of the wires and rods, since its circuit would have been along paths of low reluctance wholly within the material, should have been comparatively stable and free from disturbance by external magnetic fields of moderate intensity. The observed fact that the phenomenon was quite independent of the orientation of the specimen in the earth's magnetic field was consistent with this view. The further fact that the imposition of a strong

magnetic field along the axis substantially weakened the effect was additional confirmation, for it is well known [6] of ferro-magnetic materials that strong magnetization in a given direction reduces their susceptibility in other directions.

Of course it was inferred at once that the same stresses that brought about changes in the intensities of circular magnetization of the specimens also were causing changes in the intensities of longitudinal magnetization. A coil of insulated wire was connected to the input of the amplifier. When one of the ferro-magnetic specimens was placed along the axis of this coil so that the coil winding was approximately midway between its ends, and then was stimulated to vibration, similar sounds were heard from the speaker as with the previous arrangement. As an interesting comparison it was found with a soft steel rod specimen six feet long and one half inch in diameter that a few more than fifty turns of wire on the coil were necessary to produce a sound in the speaker of the same loudness as that obtained when the amplifier input was simply connected to the two ends of the rod. Taking account of the cross-sectional areas available to the circular and to the longitudinal magnetizations, it thus was shown that the two classes of effect were not of different orders of magnitude.

There now was prepared a specimen planned especially to emphasize the effect of circular magnetization. A soft steel tube six feet long and having an external diameter of three eighths inch and a bore diameter of one eighth inch was obtained, and midway between the two ends a small opening was cut between the outer surface and the bore. Two similar windings of insulated wire were placed, each encircling closely with four turns the longitudinal wall cross-section of the tube between one end and the mid-point. Switching arrangement was provided for connecting these two windings in series either so as to encircle the total wall cross-section in one sense, or so as to encircle the two halves in opposite senses. The tube then was placed at the axis of a six-foot long solenoid, and the entire assembly was mounted with its axis horizontal and lying in the magnetic East-West direction. Sources of direct current and of sixty cycle alternating current were available.

Demagnetization of the specimen was accomplished by passing initially strong alternating current through the solenoid and the bore windings, either successively or simultaneously, and then tapering this current off uniformly to zero value. The effectiveness of the treatment could be inferred from the following observations. The specimen was made to acquire strong residual magnetism in the longitudinal direction by passing direct current momentarily through

the solenoid winding, and its magnetized condition was verified by exploration with a magnetic compass. Then upon applying the demagnetizing cycle either to the solenoid or to the bore winding this evidence of the magnetized state would disappear.

With the halves of the bore winding in series aiding and connected to the amplifier input, tapping on the end of the demagnetized tube produced a low but distinct sound in the speaker. Upon reversing one half of the bore winding the loudness of this sound usually was slightly reduced, but occasionally was slightly increased. Following the momentary passage of direct current through the bore winding with its halves connected in either sense, the loudness of the sound from the speaker was tremendously and permanently increased. Now upon reversing one half of the bore winding the loudness of this sound always was reduced markedly—although never to so low a level as that produced by the demagnetized specimen. Also, it was noted that with the bore windings in series aiding and connected to the amplifier input and starting with the tube demagnetized, the momentary passage of direct current through the solenoid winding (thereby imposing a state of residual longitudinal magnetization upon the specimen) was followed always by a moderate but marked increase in the response to tapping.

The foregoing results established quite firmly that the effect under investigation was due largely if not wholly to variations in the intensity of circular magnetization accompanying the applications of stresses. The presence of the effect to a slight degree with a specimen which presumably was in a demagnetized state remained unexplained, since testing equipment was not available for extending the inquiry further. Several plausible explanations suggested themselves. Perhaps a specimen of ferro-magnetic material could not be demagnetized completely, or—what amounted to the same thing—perhaps the state of complete demagnetization was unstable, and was followed immediately and spontaneously by the appearance of some magnetization. Again, it might have been that the state of complete demagnetization was reasonably stable of itself, but was readily disturbed by the initial application of the mechanical stresses. It seemed reasonable to expect that any such self-magnetization would have arisen most pronouncedly along the paths of least reluctance. For the time being, it was necessary to leave this point to conjecture. Certain of the results described in the paragraph immediately preceding clearly were attributable to lack of homogeneity of the specimen.

That a simple length of iron rod should be capable of functioning as a complete alternating current generator appealed to the writers

as being novel and curious. So direct a means of converting mechanical into electrical energy should find some useful applications. As the sensitive element in a telephone transmitter it should be added to the considerable list of other devices which have been used for this and allied purposes. The following arrangement was constructed. A fine iron wire was laced back and forth between pegs located along the opposite edges of a five-inch square opening in a wooden frame so as to screen the aperture with one hundred spans of the wire all in series, evenly spaced, and parallel. A sheet of paper then was cemented to the screen and the two ends of the wire were connected to the amplifier input. This device performed as a crude telephone transmitter. It was recognized, of course, that with this simple arrangement the iron wire would undergo two complete cycles of stress for each complete cycle of the air pressure upon the diaphragm, and that mechanical bias or some equally effective means would have been necessary to eliminate this distortion. Where the vibrating element itself was of magnetic material, there was the possibility of the sound source constituting its own transmitter. For example, when the amplifier input was connected between the bridge and key-head of a steel-stringed guitar the music of this instrument was reproduced quite faithfully from the speaker. This was suggestive of the possible use of the effect in studying the vibrations and strains occurring in steel structures such as bridges.

References

1. "The Bell Telephone," American Bell Telephone Company, Boston, 1908, p. 54. The following quotation from Bell's testimony was called to the writers' attention by Mr. R. I. Caughey. *Yes. I tried such an experiment, I think in the year 1879. A continuous current from a voltaic battery was passed through a stretched wire—I think a thin steel wire—which was placed in the same circuit with an ordinary commercial hand telephone. When the wire was plucked by hand, it vibrated, producing a musical tone. The hand telephone was in another room, and I listened at the telephone while a young man was plucking the wire. I heard a musical tone from the telephone at each pluck, and could recognize, also, that the character or "timbre" of the sound produced by the vibrating wire was reproduced by the telephone at my ear.*
2. Ewing, "Magnetic Induction in Iron and Other Metals," 3d ed., 1900, Chap. IX.
3. Von Hippel and Stierstadt, *Zeitschrift fur Physik*, 69, 52 (1931).
   Von Hippel, Stierstadt, and von Auwers, *Zeitschrift fur Physik*, 72, 266 (1931).
4. Von Auwers, *Zeitschrift fur Physik*, 78, 230 (1932).
5. From the results at this preliminary stage the present writers were convinced that the electromotive forces appearing at the terminals of the vibrating wire were determined by the cycles of strain, and not by the cycles of displacement, of the specimen. It seemed to them that this fact, and this fact alone, was demonstrated by von Auwers' elaborate study of the frequency and phase relationships between the mechanical vibration and the terminal electromotive forces.
6. Ewing, Chap. IX, p. 234.

An Extension of Operational Calculus

By JOHN R. CARSON

THE Heaviside operational calculus postulates at the outset that the initial (boundary) conditions at reference time $t = 0$ are those of equilibrium; that is to say, the system is at rest when suddenly energized at time $t = 0$ by a "unit" impressed force. By unit impressed force is to be understood a force which is zero before, unity after, time $t = 0$.

In a paper published in Volume 7, 1929, of the *Philosophical Magazine*, Van der Pol briefly indicated the appropriate procedure for extending the operational calculus to cover arbitrary initial conditions. The present paper is an exposition of this generalization for a system of a finite number of degrees of freedom, followed by an application to the differential equations of the transmission line. While stated in the language of electric circuit theory, it is to be understood that the processes are generally applicable to a wide variety of problems.

We start with the canonical equations for a network of n degrees of freedom

$$\begin{aligned} z_{11}I_1 + z_{12}I_2 + \cdots + z_{1n}I_n &= E_1 \\ \cdots\cdots\cdots\cdots & \\ z_{n1}I_1 + z_{n2}I_2 + \cdots + z_{nn}I_n &= E_n \end{aligned} \tag{1}$$

where

$$z_{jk} = \left(L_{jk}\frac{d}{dt} + R_{jk} + \frac{1}{C_{jk}}\int_{-\infty}^{t} dt \right). \tag{2}$$

Now multiply the equations (1) by e^{-pt} throughout and integrate from 0 to infinity. Also let J_m and F_m denote the Laplace transforms of I_m and E_m; thus

$$\begin{aligned} J_m &= \int_0^\infty I_m e^{-pt}\,dt, \\ F_m &= \int_0^\infty E_m e^{-pt}\,dt. \end{aligned} \tag{3}$$

Now let I_m^0 and Q_m^0 denote the initial values (at time $t = 0$) of I_m and the charge Q_m in the mth mesh; also let us replace z_{jk} of (2) by

$$z_{jk} = pL_{jk} + R_{jk} + 1/pC_{jk}. \tag{4}$$

We then have, replacing (1), the system of *algebraic* equations:

$$\begin{aligned} z_{11}J_1 + z_{12}J_2 + \cdots + z_{1n}J_n &= F_1 + G_1, \\ \cdots\cdots\cdots&\cdots\cdots, \\ z_{n1}J_1 + z_{n2}J_2 + \cdots + z_{nn}J_n &= F_n + G_n. \end{aligned} \tag{5}$$

In these equations, $G_1, G_2, \cdots, G_n$ denote the following summations:

$$\begin{aligned} G_1 &= L_{11}I_1^0 - Q_1^0/C_{11}p + L_{12}I_2^0 - Q_2^0/C_{12}p + \cdots \\ &\qquad\qquad + L_{1n}I_n^0 - Q_n^0/C_{1n}p, \\ &\cdots\cdots\cdots\cdots\cdots\cdots, \\ G_n &= L_{n1}I_1^0 - Q_1^0/C_{n1}p + \cdots + L_{nn}I_n^0 - Q_n^0/C_{nn}p. \end{aligned} \tag{6}$$

The right hand sides of equations (5) are thus known in terms of the impressed forces and the specified initial values of the currents and charges. They can therefore be solved in the usual manner for $J_1, \cdots, J_n$. Thus

$$J_m = \frac{F_1 + G_1}{Z_{m1}} + \frac{F_2 + G_2}{Z_{m2}} + \cdots + \frac{F_n + G_n}{Z_{mn}}. \tag{7}$$

Having thus determined $J_1, \cdots, J_n$ as functions of p, $I_1, \cdots, I_n$ are determined as functions of time by the Laplace integral equation:

$$J_m(p) = \int_0^\infty I_m(t)e^{-pt}\,dt, \qquad p_R > c, \tag{8}$$

which completes the formal solution of the problem. Note that if $G_1 = G_2 = \cdots = G_n = 0$, the solution reduces to the usual form.

Equations (6) for $G_1, \cdots, G_n$, may be written in a compact and elegant form as follows: Let

$$\begin{aligned} T &= \tfrac{1}{2}\sum\sum L_{jk}I_jI_k, \\ U &= \tfrac{1}{2}\sum\sum \frac{1}{C_{jk}} Q_jQ_k. \end{aligned} \tag{9}$$

T is then the kinetic or magnetic energy stored in the network and U is the corresponding potential or electric energy. Then

$$G_m = \left(\frac{\partial T}{\partial I_m}\right)_{t=0} - \frac{1}{p}\left(\frac{\partial U}{\partial Q_m}\right)_{t=0}, \; m = 1, 2, \cdots, n. \tag{10}$$

The foregoing solution is compact, elegant and formally complete. In practical applications to networks of many degrees of freedom it may well present formidable difficulties in computation and interpre-

tation. This, however, is merely answerable to the complexity of the physical problem, and no simpler general solution can possibly exist.

The foregoing method when applied to the differential equations of the transmission line, leads to the following differential equations

$$
\begin{aligned}
(Lp + R)J &= -\frac{\partial}{\partial x}\Phi + LI^0, \\
(Cp + G)\Phi &= -\frac{\partial}{\partial x}J + CV^0.
\end{aligned}
\tag{11}
$$

Here J and Φ are Laplace transforms of the current I and voltage V and I^0, V^0 are the initial values of I and V at reference time $t = 0$. J, Φ, I^0, V^0 are functions of x but of course independent of t.

The formal solution of equations (11) is as follows: write

$$
\begin{aligned}
Lp + R &= Z(p) = Z, \\
Cp + G &= Y(p) = Y, \\
\sqrt{ZY} = \gamma, &\qquad \sqrt{Z/Y} = K.
\end{aligned}
\tag{12}
$$

Also

$$
LI^0 - \frac{C}{Y}\frac{\partial}{\partial x}V^0 = F(x) = F.
$$

Then

$$
\begin{aligned}
J = e^{-\gamma x}&\left\{A + \frac{1}{2K}\int^x dyF(y)e^{\gamma y}\right\} \\
- e^{\gamma x}&\left\{B + \frac{1}{2K}\int^x dyF(y)e^{-\gamma y}\right\},
\end{aligned}
\tag{13}
$$

$$
\Phi = -\frac{K}{\gamma}\frac{\partial}{\partial x}J + \frac{C}{Y}V^0. \tag{14}
$$

A and B are constants of integration determined by the relations between J and Φ at the physical terminals of the line.

Determination of the Corrosion Behavior of Painted Iron and the Inhibitive Action of Paints *

By R. M. BURNS and H. E. HARING

THE value of paints for the protection of metal surfaces depends upon their effectiveness as physical barriers against the corrosive elements of the surrounding environment and upon the electrochemical activity of the primer pigments in rendering the surfaces passive. Physical testing methods have been developed which furnish valuable information concerning the quality and rate of aging of paint films.[1] There is, however, an obvious need for direct methods of determining the condition and behavior of the metal surface beneath the paint film, the rate of penetration of corrosive agents through the film, and the mechanism of the inhibitive action afforded by the film. The present paper describes an electrochemical method of obtaining this information.

It is well established that the process of corrosion in the presence of moisture is electrolytic in character—that it occurs by means of the operation of small galvanic cells at the surface of the metal. It should be possible, therefore, to determine the corrosion behavior of a metal by measuring the electrical characteristics of these individual corrosion cells; but the multiplicity and minute size of the anodic and cathodic areas makes such measurements impracticable. However, it is readily possible to determine the resultant of all of the polarized potentials of the anodic and cathodic areas on the metal surface, and to follow the change in this potential (of the electrode as a whole) with time.[2] Experience in this and other laboratories has demonstrated that time-potential curves obtained in this manner indicate the corrosion behavior of a metal and the state of its surface.[3] In general, it has been found that if the potential of a metal becomes more electropositive (more noble) with time, the formation of a protective film and a retardation or cessation of corrosion is taking place, while conversely, if the potential becomes more negative, solution of a protective film and acceleration of corrosion is indicated.

* Digest of a paper presented before the Spring Meeting of the Electrochemical Society at Cincinnati, Ohio, April 23–25, 1936, and published subsequently in volume 69 of its *Transactions*.

[1] Schuh, *Ind. Eng. Chem.*, 23, 1346 (1931).

[2] Burns, *Bell System Tech. Jour.*, 15, 20 (1936).

[3] May, *Jour. Inst. Met.*, 40, 141 (1928); Bannister and Evans, *Jour. Chem. Soc.* (June 1930).

The corrosion of iron can be predicted by proper interpretation of such time-potential curves. The fact that the iron is painted should not alter this conclusion, and therefore the method has been applied to a study of painted iron and primer pigments.

The customary procedure for determining time-potential curves was followed with the exception that the potentiometer ordinarily employed was replaced by a vacuum tube electrometer when measurements were made of the potential of painted iron or of iron in other media of high electrical resistance. This instrument, slightly modified to take advantage of certain improvements which have been made in compensated single tube circuits, is described elsewhere.[4]

Common usage has defined iron which is corroding as "active" and iron which is not corroding as "passive." In order to obtain a background of information which might serve as a guide in the study of painted iron and paint pigments, a series of time-potential curves depicting iron and steel in the active and passive states was determined. The results are shown in Fig. 1.

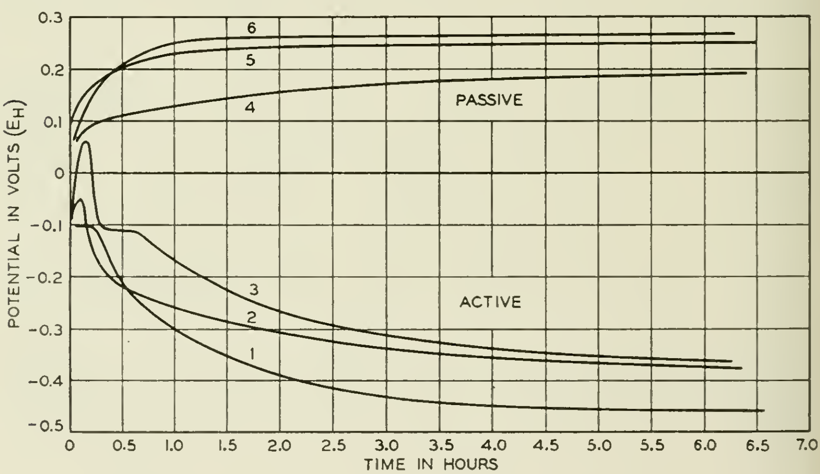


Fig. 1—Time-potential curves for active and passive iron.

Active (Corrosion)
1. Iron in tap-water.
2. Iron in 0.01N NaCl.
3. Uncleaned iron in tap-water.

Passive (No Corrosion)
4. Stainless steel in tap-water.
5. Iron in 0.01N $K_2Cr_2O_7$.
6. Uncleaned stainless steel in tap-water.

[4] Compton and Haring, *Trans. Electrochem. Soc.*, 62, 345 (1932); D. B. Penick, *Rev. Sci. Inst.*, 6, 115 (1935) and *Bell Laboratories Record*, 14, 74 (1935).

It will be noted that the potentials of the test electrodes are initially quite similar, but diverge with time and form two distinct groups of curves, which ultimately become separated by about 0.7 volt. It was observed that invariably electrodes did not corrode if their potentials became more electropositive (more noble) over a long period of time, while, on the other hand, marked corrosion accompanied a negative trend of potential. A state of equilibrium was reached ultimately by the passive electrodes between 0.25 and 0.30 volt, and by the active electrodes between −0.40 and −0.45 volt.

Red oxide and red lead paints were selected for study because practical experience indicates that they are representative of the two types of protective paint, viz., (1) those which protect merely because they serve as physical barriers, and (2) those which exert a chemical inhibiting action as well.

The test electrodes were commercial iron, of high purity, in the form of 1/8 inch rods. The pigments were technical grades of red oxide (Fe_2O_3) and red lead (Pb_3O_4) of high quality. Raw linseed oil and a lead-cobalt dryer were used in the preparation of all of the paints, which were formulated and compounded in the customary manner. Approximately 20 per cent of a flexible type varnish and 10 per cent of blown linseed oil were incorporated with raw linseed oil to form the vehicle in one of the red oxide paints.

As a rule, the primary purpose of a protective paint is to shield iron from the corrosive action of water and water vapor. Total immersion is an extreme condition, but a condition to which all such paints are frequently subjected. For this reason, and also in order to speed up possible reactions and save time, all of the potential measurements on painted iron recorded in this paper were made on submerged specimens. Similar measurements on painted iron exposed to the atmosphere are equally possible and can be made without disturbing service conditions. It is planned to extend this study to include such measurements.

The time-potential curves obtained in the study of primers are presented in Fig. 2. There are included for reference typical curves (6 and 7) for iron in the active and passive states, and a curve (curve 5) for iron coated with a dried film of linseed oil. It will be noted that the linseed oil coated electrode behaved in much the same manner as bare, active iron, except that a much longer time was required for the potential changes to take place. Several days elapsed before the potential reached the equilibrium value attained by bare iron in a few hours, and at this point rust was clearly visible.

The potential of iron painted with red oxide primer (curve 1), immediately after immersion in water, was approximately 0.33 volt.

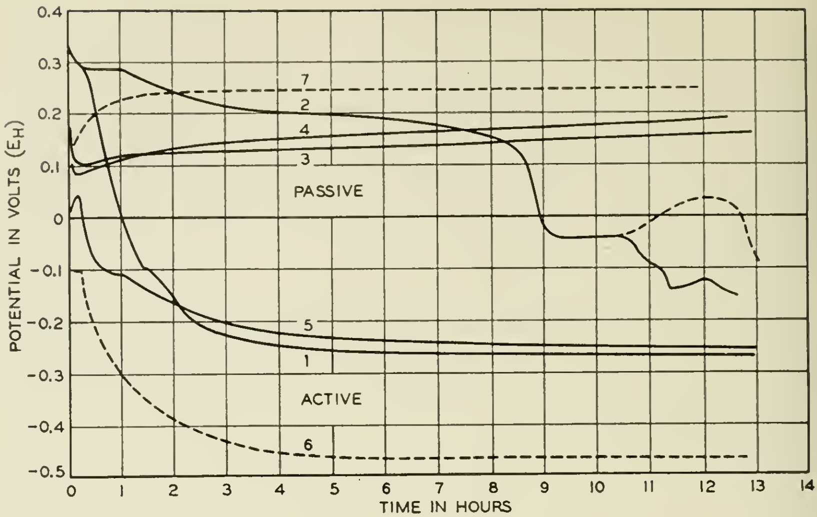


Fig. 2—Time-potential curves for painted iron in water.

Red Oxide Primer
1. Linseed oil vehicle.
2. Linseed oil plus varnish.

Red Lead Primer
3. No visible pores.
4. Visible pores.

Reference Curves
5. Dry linseed oil film.
6. Iron in tap-water.
7. Iron in 0.01N $K_2Cr_2O_7$.

In other words, the metal was passive initially and, judging from the behavior of the curve, it tended to remain so for a period of about twenty minutes. By the end of this time, water appears to have permeated the paint film in sufficient quantity to induce the active corrosion indicated by the sharp drop in potential which followed. From this time on, the curve is similar to that for the linseed oil coated electrode. Ultimate equilibrium required a much longer time, however. After several days' immersion, the paint film was removed and the iron was found to be corroded. The very gradual slope of the curves for active painted electrodes, after their sharp breaks from the passive region, may be attributed to the action of the films as partial barriers to moisture.

The impermeability of primers and their adherence to iron are known to be increased by the addition of varnish. Accordingly, red oxide

primer containing a moderate amount of varnish was subjected to test. The results are represented by curve 2 in Fig. 2.

This curve is of especial significance, not only because its characteristics are so much more pronounced than those of curve 1, but also because it furnishes an explanation for the divergence of results which have been obtained with red oxide primers in practice. The marked increase in the length of time required for this curve to pass through its various phases as compared to curve 1, is evidence of the water-excluding effect of the added varnish. The period of definite passivity has been extended to at least three times its former length, and the momentary halt in curve 1 at approximately -0.09 volt has been prolonged to an hour and a half at a slightly more positive potential. Curve 2 continues somewhat erratically in the active direction after the half-way halt in its course. Examination of the iron after several days' immersion revealed corrosion. The broken line represents, in days rather than hours, the quite different behavior of a duplicate specimen, which prior to this time had acted similarly except for the fact that it had required a somewhat longer time to pass through its various phases. Apparently a slightly less permeable paint film made it possible for the corrosive action to be stifled, temporarily at least.

Alternations of corrosive attack and film formation were observed generally when iron corroded in contact with red oxide pigments and primers. In a relatively dry atmosphere there is no doubt that iron painted with red oxide is maintained in a passive condition, but exposure to excess moisture must result eventually in active corrosion. Since, then, the protective value of red oxide primers is dependent primarily upon their ability to exclude moisture, they must be classed as physical inhibitors of corrosion.

The corrosion behavior of iron painted with red lead is clearly indicated in Fig. 2 by curves 3 and 4. The presence of a few small pores or imperfections in the paint film on one of the test electrodes did not materially affect the results. The initial potentials were somewhat lower than was the case for red oxide, and the initial trend of the curves was in the active direction, but a reversal soon took place and the iron became definitely and permanently passive. An equilibrium potential of approximately 0.25 volt was attained. Inspection of the iron after several weeks' immersion failed to reveal any sign of corrosion.

Red lead primer continues to inhibit corrosion even after moisture has fully penetrated the film. On the basis of its action, both as pigment and primer, red lead must be classed as a chemical inhibitor of corrosion. The reason for its passivating action is a disputed question. Paint chemists have inclined to the view that a highly protective lead

soap is formed, but this theory becomes untenable in view of evidence that red lead pigment alone passivates iron in much the same manner as red lead primer, and that even water solutions of the pigment have an effect. Other theories are that the iron is rendered passive by the alkalinity of the red lead, or because it is an oxidizing agent. In all probability, both of these factors are involved.

The ease with which it has been found possible to make potential measurements on painted iron with the aid of the vacuum tube electrometer, suggests the application of the time-potential method of study to the determination of the corrosion behavior of iron encased in concrete or buried underground or immersed in oil or other highly resistant media. Field study would be facilitated by substitution of a vacuum tube voltmeter for the electrometer.

The fact that there is a potential difference of at least 0.5 volt between the active and passive states of iron suggests a rapid potentiometric method for the determination of the permeability of all types of organic coatings. The time-potential curve for an iron electrode coated with the organic material and immersed in a salt solution, for example, would break sharply at the moment penetration was attained.

Abstracts of Technical Articles from Bell System Sources

The Orientation of Crystals in Silicon Iron.[1] RICHARD M. BOZORTH. X-ray examination of silicon iron prepared by N. P. Goss shows that the component crystals are oriented so that a [001] direction is parallel to the direction of rolling and a (110) plane lies in the rolling plane. This is contrary to the result reported by Goss in his paper "New Development in Electrical Strip Steels Characterized by Fine Grain Structure Approaching the Properties of a Single Crystal," published in *Transactions of the American Society for Metals*, Volume 23, June, 1935, page 511. The differences in the magnetic properties in different directions in the sheet are explained in terms of the properties of the single crystals.

Eddy Currents in Composite Laminations.[2] E. PETERSON and L. R. WRATHALL. The familiar theory of eddy current shielding leads to an expression for the impedance of a ferromagnetic core inductance coil in terms of the initial permeability and resistivity of the core material, the core geometry, and the measuring frequency. Measurements on a number of different core materials over a wide frequency range have revealed sizeable deviations from the theory in some cases. The discrepancies are especially marked in some specimens of chromium permalloy, the measured inductance over a certain frequency range being of the order of one tenth that specified by the theory.

It appears that discrepancies arise when the laminations are not homogeneous, a condition contrary to an assumption of the simple theory. The inhomogeneity takes the form of a thin surface layer which has a permeability much less than that of the interior. By etching off these surface layers, the initial permeability is increased, and discrepancies between the measured variations of impedance with frequency and those calculated for a homogeneous sheet are removed almost completely.

The theory has been extended to take account of the surface layers, and agrees well with measurements on the original unetched laminations when plausible assumptions are made regarding the properties of the surface layer.

[1] *Transactions, Amer. Soc. for Metals*, December, 1935.
[2] *Proc. I. R. E.*, February, 1936.

Applications of X-Ray Photography in Industrial Development Work.[3] J. R. TOWNSEND and L. E. ABBOTT. The fundamentals of X-ray technic as applied in developmental work of the telephone industry are outlined. A brief description is presented of the physics of X-rays, of the methods used to produce usable X-ray radiation, and of a typical industrial X-ray laboratory. Results are given of investigations at the Bell Telephone Laboratories to determine the sensitivity of X-ray methods of revealing internal defects in metals. Numerous examples illustrating the application of X-rays in telephone work are included, as well as a description of the use of gamma rays for industrial application.

Principles of Measurements of Room Acoustics.[4] E. C. WENTE. The acoustic characteristics of a room can in great part be evaluated from a knowledge of the rate with which sound in the room dies down when emission from the source ceases. The physical principles underlying the relationship are briefly discussed. It is shown by specific examples that we can obtain valuable additional information about acoustics of a room by recording the sound level at one or more points in the room when the frequency of the sound is continuously varied.

Visual Accompaniment.[5] R. WOLF. The principles of producing "Visual Accompaniments" to musical renditions for the theater are briefly described, as follows: (1) natural scenes for portraying the "musical mood" of the musical composition; (2) the changing and blending of beautiful paintings to interpret the mood, known as the Savage Method; and (3) the use of abstract color forms as a means of interpretation. The technic followed in applying the two latter methods is described in detail.

[3] Presented at the Fall 1934 mtg. of S. M. P. E., New York, N. Y.; published in somewhat condensed form in *Metal Progress*, February, 1936, under the title "Some Applications of X-Rays to Industrial Problems."
[4] *Jour. S. M. P. E.*, February, 1936.
[5] *Jour. S. M. P. E.*, February, 1936.

Contributors to this Issue

R. M. Burns, A.B., University of Colorado, 1915; A.M., 1916; Ph.D., Princeton University, 1921; Instructor, University of Colorado, 1916–17. Second Lieutenant, Chemical Warfare Service, U. S. Army, 1918–19. Research chemist, Barrett Company, 1921–22. Western Electric Company, 1922–25. Bell Telephone Laboratories, 1925–; Assistant Chemical Director, 1931–. Dr. Burns' work has been largely in the electrochemical field and particularly on the subject of the corrosion of metals and its prevention.

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912. American Telephone and Telegraph Company, 1914–34; Bell Telephone Laboratories, 1934–. As Transmission Theory Engineer for the American Telephone and Telegraph Company and later for the Laboratories, Mr. Carson has made substantial contributions to electric circuit and transmission theory and has published extensively on these subjects. He is now a research mathematician.

Carl J. Christensen, B.S., Brigham Young University, 1923; M.S., University of Wisconsin, 1925; Ph.D., University of California, 1929. Instructor in Chemistry and Physics, Brigham Young University, 1923–24 and 1925–27. Bell Telephone Laboratories, 1929–. Dr. Christensen has been engaged in research in connection with microphone carbon.

H. E. Curtis, S.B. and S.M. in Electrical Engineering, Massachusetts Institute of Technology, 1929. American Telephone and Telegraph Company, Department of Development and Research, 1929–34; Bell Telephone Laboratories, 1934–. Mr. Curtis has been engaged in high-frequency transmission problems as related to carrier systems.

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917. Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

E. I. Green, A.B. Westminster College, 1915; University of Chicago, 1915–16; Member of faculty, Westminster College, 1916–17; U. S. Army, 1917–19 (Captain, Infantry); B.S., Harvard University, 1921. Department of Development and Research, American Telephone and Telegraph Company, 1921–34; Bell Telephone Laboratories, 1934–. Mr. Green has been engaged principally in work on multiplex wire transmission systems.

H. E. Haring, B.S., Franklin and Marshall College, 1916; M.A., Princeton University, 1917. Assistant Chemist, Ordnance Department, U. S. Army, 1917–19; Associate Chemist, U. S. Bureau of Standards, 1919–28; Electrochemist, Victor Talking Machine Company, 1928–29; Bell Telephone Laboratories, 1929–. Since 1919 Mr. Haring has been engaged in electrochemical research in connection with storage batteries and other electrochemical apparatus, electrodeposition, and corrosion.

H. P. Lawther, Jr., B.A., University of Texas, 1912; M.A., Harvard, 1913; Ph.D., Harvard, 1916. Southwestern Bell Telephone Company, 1919–. Transmission and Protection Engineer, Texas Area, 1928–32; General Transmission and Protection Engineer, St. Louis, Missouri, 1932–.

F. A. Leibe, M.E., Stevens Institute of Technology, 1922. Westinghouse Lamp Company, 1922–25; American Telephone and Telegraph Company, Department of Development and Research, 1925–34; Bell Telephone Laboratories, 1934–. Mr. Leibe has been engaged in transmission development work on carrier telephone systems.

Sallie Pero Mead, A.B., Barnard College, 1913; M.A., Columbia University, 1914. American Telephone and Telegraph Company, Engineering Department, 1915–19; Department of Development and Research, 1919–34. Bell Telephone Laboratories, 1934–. Mrs. Mead's work has been of a mathematical character relating to telephone transmission.

G. L. Pearson, A.B., Willamette University, 1926; M.A., Stanford University, 1929. Bell Telephone Laboratories, 1929–. Mr. Pearson has been engaged in a study of noise in electric circuits.

W. O. Pennell, B.Sc., Massachusetts Institute of Technology, 1896; Instructor, Lafayette College, 1896–98. Equipment and Traffic Engineer, Bell Telephone Company of Philadelphia, 1898–1902; Engineering Department, American Telephone and Telegraph

Company, 1902–03; Chief Engineer, Missouri and Kansas Telephone Company, 1903–12; Southwestern Bell Telephone Company, St. Louis, Missouri, 1912–, and for the last eighteen years its Chief Engineer. Mr. Pennell has from time to time contributed articles to various technical publications.

Arthur C. Peterson, Jr., B.S. in Electrical Engineering, University of Washington, 1928. American Telephone and Telegraph Company, Department of Development and Research, 1928–34; Bell Telephone Laboratories, 1934–. Mr. Peterson's work has been primarily concerned with short-wave radio transmission development.

Ralph Kimball Potter, B.S., Whitman College, 1917; Artillery Corps, U. S. Army, 1917–19; E.E., Columbia University, 1923. American Telephone and Telegraph Company, Department of Development and Research, 1923–34; Bell Telephone Laboratories, 1934–. Mr. Potter has been engaged in short-wave radio transmission development work.

S. A. Schelkunoff, B.A., M.A. in Mathematics, The State College of Washington, 1923; Ph.D. in Mathematics, Columbia University, 1928. Engineering Department, Western Electric Company, 1923–25. Bell Telephone Laboratories, 1925–26. Department of Mathematics, State College of Washington, 1926–29. Bell Telephone Laboratories, 1929–. Dr. Schelkunoff has been engaged in mathematical research, especially in the field of electromagnetic theory.

George C. Southworth, B.S., Grove City College, 1914; Sc.D. (Hon.), 1931; Ph.D., Yale University, 1923. Assistant Physicist, Bureau of Standards, 1917–18; Instructor, Yale University, 1918–23. Editorial staff of *Bell System Technical Journal*, American Telephone and Telegraph Company, 1923–24; Department of Development and Research, 1924–34; Research Department, Bell Telephone Laboratories, 1934–. Dr. Southworth's work in the Bell System has been concerned chiefly with the development of short-wave radio telephony. He is the author of several papers on very high-frequency electrical phenomena.

VOLUME XV JULY, 1936 NUMBER 3

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each. The foreign postage is 35 cents per year or 9 cents per copy.

The Bell System Technical Journal

Vol. XV *July, 1936* *No. 3*

A Laplacian Expansion for Hermitian-Laplace Functions of High Order*

By E. C. MOLINA

Among the wide variety of practical and theoretical problems confronting the telephone engineer, there is a surprisingly large number to whose solution mathematics has made notable contribution. In his kit of mathematical tools the theory of probability is a frequently used and most effective instrument. This theory of probability contains a large number of theorems, a large number of functions, which permit of application to telephony. Among these is a particular tool, a particular group of mathematical functions known as the "Hermitian Functions," each of which is identified by a number called its "order." These mathematical functions or relations have no practical utility until the variables in the equation can be assigned numerical values and the resultant numerical value of the function calculated. Tables of the numerical values of Hermitian functions of low order exist; for example, Glover's Tables of Applied Mathematics cover the ground for those of the first eight orders. But tables for the functions of higher order are still a desideratum. This paper presents an expansion by means of which the evaluation of a high order function can be readily accomplished with a considerable degree of accuracy.

The development of the expansion is prefaced by some remarks on the early history of the Hermitian functions and the relation of this history to modern theoretical physics.

I

AMONG contributions made by Laplace to the domain of pure and applied mathematics, two of great practical value are:

(*a*) His method of evaluating definite integrals [1] whose integrands involve factors raised to high powers;

(*b*) The pair of orthogonal polynomial functions [2] which he defined by the following Equations (1) and (2)

$$(1)\qquad [(2n)!\sqrt{\pi}/2^{2n}n!]U_n(u) = \int_{-\infty}^{\infty} e^{-x^2}(x-iu)^{2n}dx = 2e^{u^2}\int_0^{\infty} e^{-x^2}x^{2n}\cos(2ux)dx;$$

$$(2)\qquad [(2n+1)!\sqrt{\pi}/2^{2n}n!]U_n'(u) = i\int_{-\infty}^{\infty} e^{-x^2}(x-iu)^{2n+1}dx = 2e^{u^2}\int_0^{\infty} e^{-x^2}x^{2n+1}\sin(2ux)dx.$$

* Presented at International Congress of Mathematicians, Oslo, Norway, July 13–18, 1936.

These polynomial functions formed the coefficients in a series satisfying the partial differential equation

$$\frac{\partial U}{\partial r'} = 2U + 2u\frac{\partial U}{\partial u} + \frac{\partial^2 U}{\partial u^2}$$

to which Laplace reduced the solution of the following ball problem: [3]

Consider two urns A and B each containing n balls and suppose that of the total number of balls, $2n$, as many are white as black. Conceive that we draw simultaneously a ball from each of the urns, and that then we place in each urn the ball drawn from the other. Suppose that we repeat this operation any number, r, of times, each time shaking well the urns in order that the balls be thoroughly mixed; and let us find the probability that after the r operations the number of white balls in urn A be x.

Under the caption "The Statistical Meaning of Irreversibility" Lotka [4] has pointed out the significance of Laplace's ball problem in the modern kinetic theory of matter. Moreover, Hostinsky [5] has shown the bearing of the same problem on the theory of Brownian movements and said "In effect, the partial differential equation obtained by Laplace has been refound by Smoluchowski."

To avoid confusion with the Laplace functions which one encounters in spherical harmonic analysis, the functions defined by Equations (1) and (2) are herein designated as Hermitian-Laplace functions. Such a designation is justified by the Equations (3) and (4) derived in the next paragraph.

II

We also find in Laplace [6]

$$I_n(u)^* = \int_0^\infty e^{-x^2}x^{2n}\cos(2ux)dx = \frac{(-1)^n\sqrt{\pi}}{2^{2n+1}}\left(\frac{d^{2n}e^{-u^2}}{du^{2n}}\right),$$

$$I_n'(u) = \int_0^\infty e^{-x^2}x^{2n+1}\sin(2ux)dx = \frac{(-1)^{n+1}\sqrt{\pi}}{2^{2n+2}}\frac{(d^{2n+1}e^{-u^2})}{du^{2n+1}}.$$

Comparing these Laplacian expressions for the definite integrals $I_n(u)$ and $I_n'(u)$ with the Equations (1) and (2) we see immediately that

$$(3) \qquad U_n(u) = (-1)^n[n!/(2n)!]H_{2n}(u),$$

$$(4) \qquad U_n'(u) = (-1)^{n+1}[n!/2(2n+1)!]H_{2n+1}(u),$$

where H_{2n} and H_{2n+1} are the original Hermite polynomials [7] of order $2n$ and $2n+1$, respectively. These equations connecting the Her-

* The symbols $I_n(u)$ and $I_n'(u)$ are introduced here as convenient abbreviations for the integrals to which they are equated; these symbols do not appear in Laplace.

mite with the Laplace polynomials have been presented in an earlier paper.[8]

Appell and Feriet, Arne Fisher, T. C. Fry, H. L. Rietz and others base their definitions of the Hermite polynomials on $e^{-x^2/2}$ instead of e^{-x^2}. We shall write $A_n(u)$ for the nth polynomial as defined by these authors, reserving $H_n(u)$ to symbolize the Hermitian polynomial as defined in his paper of 1864. Thus, in what follows,

$$A_n(x) = (-1)^n e^{x^2/2}(d^n e^{-x^2/2}/dx^n), \qquad H_n(u) = (-1\sqrt{2})^n A_n(u\sqrt{\ }$$

III

Laplacian expansions * for the U, H, and A polynomials follow immediately from those obtainable by applying to the integrals $I_n(u)$ and $I_n'(u)$ his method of evaluating definite integrals whose integrands embrace factors raised to high powers. As will be shown in Part IV of this paper, we have

$$I_n(u)/[\sqrt{\pi}(Y\sqrt{N})^N] = [S\cos(u\sqrt{2N})] + [S'\sin(u\sqrt{2N}], \qquad N = 2n,$$

$$I_n'(u)/[\sqrt{\pi}(Y\sqrt{N})^N] = [S\sin(u\sqrt{2N})] - [S'\cos(u\sqrt{2N}], N = 2n + 1,$$

where $Y = (xe^{-x^2})$ for $x = X = 1/\sqrt{2}$ and

$$S = \sum_{s=0}^{\infty}\left(\frac{-1}{4N}\right)^s [u^{-(2s+1)}K_{2s}],$$

$$S' = \left(\frac{1}{2\sqrt{N}}\right)\sum_{s=1}^{\infty}\left(\frac{-1}{4N}\right)^{(s-1)}[u^{-2s}K_{2s-1}].$$

The explicit expressions for K_0, K_2, K_4 and K_1, K_3, K_5 are given in Section V of this paper. The desired expansions are then given by the equations:

$$\begin{aligned}
e^{u^2}I_n(u)/\sqrt{\pi} &= U_n(u)[(2n)!/2^{2n+1}n!] \\
&= H_{2n}(u)[(-1)^n/2^{2n+1}] \\
&= A_{2n}(u\sqrt{2})[(-1)^n/2^{n+1}], \\
e^{u^2}I_n'(u)/\sqrt{\pi} &= U_n'(u)\,[(2n+1)!/2^{2n+1}n!] \\
&= H_{2n+1}(u)[(-1)^{n+1}/2^{2n+2}] \\
&= A_{2n+1}(u\sqrt{2})[(-1)^n/2^{n+1}\sqrt{2}].
\end{aligned}$$

The numerical results shown below in Table I indicate the efficacy

* It may be of interest to compare the expansions presented in this paper with the asymptotic forms of the Hermite functions given by N. Schwid[9] and by M. Plancherel and M. Rotach.[10]

TABLE I

N	$u\sqrt{2} = x$	True Value of $A_N(x)$	$[{}_sA_N(x) - A_N(x)]/A_N(x)$		
			$s = 1$	$s = 2$	$s = 3$
9	0.1	9.32438×10^1	0.0089	0.0000	0.0000
	0.5	3.26533×10^2	0.0084	0.0001	0.0000
	1	2.80000×10^1	0.0263	− 0.0010	− 0.0002
	2	-1.90000×10^2	0.0002	0.0018	0.0001
	3	1.62000×10^3	− 0.0474	− 0.0027	0.0002
	4	-1.74680×10^4	− 0.0283	− 0.0082	− 0.0027
10	0.1	-8.98064×10^2	0.0084	− 0.0001	0.0000
	0.5	4.90439×10^1	− 0.0027	0.0013	0.0000
	1	1.21600×10^3	0.0046	0.0003	0.0000
	2	-2.62100×10^3	− 0.0147	0.0002	0.0001
	3	9.50400×10^3	− 0.0436	− 0.0044	− 0.0004
	4	-5.18090×10^4	0.0445	0.0013	− 0.0018
15	0.1	-1.98001×10^5	0.0054	0.0000	0.0000
	0.5	-5.05845×10^5	0.0052	0.0000	0.0000
	1	4.69456×10^5	0.0022	0.0002	0.0000
	2	-1.41980×10^6	− 0.0102	0.0000	0.0000
	3	4.38955×10^6	− 0.0284	− 0.0017	− 0.0001
	4	-1.85644×10^7	− 0.0338	− 0.0041	− 0.0006
20	0.1	5.90233×10^8	0.0042	0.0000	0.0000
	0.5	-4.45178×10^8	0.0035	0.0000	0.0000
	1	-1.61935×10^8	0.0046	0.0000	0.0000
	2	-1.62882×10^9	− 0.0081	0.0000	0.0000
	3	4.60718×10^9	− 0.0212	− 0.0009	0.0000
	4	8.53219×10^9	0.1241	0.0068	0.0000

of the Laplacian expansion as applied to the evaluation of $A_n(x)$, $x = u\sqrt{2}$, for values of x ranging from 0.1 to 4 and for $N = 9$, 15, 10 and 20, respectively.

Designating by ${}_sA_N(x)$ the approximate value obtained for $A_N(x)$ when one takes into account the first s terms in each of the two series S, S', the last three columns of the table show the proportional errors incurred when $s = 1$, 2 and 3, respectively. It will be noted that for $N = 10$ and $x = 4$ the second approximation is closer than the third; this situation will occasion no surprise if it is recalled that to obtain the best results the natural order of the terms may have to be altered when, for example, one expresses a term of the binomial expansion in a series of Hermite functions.

For the convenience of one who wishes to calculate $A_N(x)$ for values of N other than 9, 10, 15 and 20, there are given in Table II the values of $u^{-(m+1)}K_m$ for $m = 0, 1, 2, 3, 4, 5$ and those values of u covered by Table I.

I am indebted to Miss E. V. Wyckoff of Bell Telephone Labora-

tories, for the computations involved in the preparations of Tables I and II.

TABLE II

$u\sqrt{2}$	$u^{-1}K_0$	$u^{-3}K_2$	$u^{-5}K_4$
0.1	0.7053412	0.234229	0.034490
0.5	0.6642654	0.198970	− 0.069716
1	0.5506953	0.0936947	− 0.273541
2	0.2601300	− 0.158968	− 0.018172
3	0.07452849	− 0.121885	0.514828
4	0.01295111	0.0244632	− 0.078479
	$u^{-2}K_1$	$u^{-4}K_3$	$u^{-6}K_5$
0.1	− 0.03520828	0.00602953	0.091905
0.5	− 0.1591469	0.0450683	0.40870
1	− 0.2294564	0.150921	0.48658
2	− 0.08671002	0.255634	− 0.78050
3	0.05589638	− 0.104689	− 0.35023
4	0.04317037	− 0.162097	0.98512

IV

A simple change of variable gives

$$(5) \quad I_n(u) = (\sqrt{N})^{N+1}\int_0^\infty (e^{-x^2}x)^N \cos(x2u\sqrt{N})dx, \qquad N = 2n,$$

$$(6) \quad I_n'(u) = (\sqrt{N})^{N+1}\int_0^\infty (e^{-x^2}x)^N \sin(x2u\sqrt{N})dx, \qquad N = 2n+1.$$

Set $y(x) = e^{-x^2}x$, and note that $dy/dx = 0$ for $x = X = 1/\sqrt{2}$. Now set $Y = y(X)$,

$$[g(x)]^2 = (\log Y - \log y)/(x-X)^2$$
$$= \frac{1}{X^2}\left[1 - \frac{1}{3}\left(\frac{x-X}{X}\right) + \frac{1}{4}\left(\frac{x-X}{X}\right)^2 - \frac{1}{5}\left(\frac{x-X}{X}\right)^3 + \cdots\right],$$

(7), $t = (x - X)g(x)$.

These transformations give

$$I_n(u) = (\sqrt{N})^{N+1}Y^N\int_{-\infty}^{\infty} e^{-Nt^2}(dx/dt)\cos(x2u\sqrt{N})dt.$$

By (7) and the Lagrange-Laplace expansion for a function of x in powers of t we obtain

$$I_n(u) = (\sqrt{N})^{N+1}Y^N\sum_{m=0}^{\infty}\int_{-\infty}^{\infty} e^{-Nt^2}[t^{2m}A_{2m}/(2m)!]dt$$

or

$$I_n(u)/\sqrt{\pi}(Y\sqrt{N})^N = \sum_{m=0}^{\infty} (1/2\sqrt{N})^{2m}(A_{2m}/m!),$$

where, writing D for the differential operator d/dx,

$$A_{2m} = [D_x^{2m} g^{-(2m+1)} \cos (x2u\sqrt{N})]_{x=X}$$

or, by the Leibnitz theorem for the product of two functions,

$$A_{2m} = \sum_{r=0}^{2m} \binom{2m}{r} (2u\sqrt{N})^r \cos (u\sqrt{2N} + r\pi/2)[D_x^{2m-r} g^{-(2m+1)}]_{x=X}$$

and, therefore,

$$\frac{A_{2m}}{m!} = \cos (u\sqrt{2N}) \sum_{r=0}^{m} \binom{2m}{2r} \frac{(-1)^r (u2\sqrt{N})^{2r}}{m!} [D_x^{2m-2r} g^{-(2m+1)}]_{x=X}$$

$$- \sin (u\sqrt{2N}) \sum_{r=0}^{m-1} \binom{2m}{2r+1} \frac{(-1)^r (u2\sqrt{N})^{2r+1}}{m!} [D_x^{2m-2r-1} g^{-(2m+1)}]_{x=X},$$

on separating the even and odd terms in r. Now setting $m - r = s$ and summing with reference to s and r, instead of m and r, gives

$$\sum_{m=0}^{\infty} (1/2\sqrt{N})^{2m} A_{2m}/m!$$

$$= \cos (u\sqrt{2N}) \sum_{s=0}^{\infty} \left(\frac{1}{2\sqrt{N}}\right)^{2s} \frac{1}{(2s)!} \sum_{r=0}^{\infty} \frac{(2r + 2s)!}{(2r)!} \frac{(-u^2)^r}{(r+s)!}$$

$$\times [D_x^{2s} g^{-(2r+2s+1)}]_{x=X} - \sin (u\sqrt{2N}) \sum_{s=1}^{\infty} \left(\frac{1}{2\sqrt{N}}\right)^{2s-1} \frac{1}{(2s-1)!}$$

$$\times \sum_{r=0}^{\infty} \frac{(2r+2s)!}{(2r+1)!} \frac{(-1)^r u^{2r+1}}{(r+s)!} [D_x^{2s-1} g^{-(2r+2s+1)}]_{x=X}.$$

But, writing $u/g = v$, we have

$$\sum_{r=0}^{\infty} \left(\frac{(2r+2s)!}{(2r)!}\right) \frac{(-1)^r u^{2r}}{(r+s)!} [D_x^{2s} g^{-(2r+2s+1)}]$$

$$= (-1)^s u^{-(2s+1)} \left[D_x^{2s} v^{2s+1} \sum_{r=0}^{\infty} \frac{(-1)^{r+s} v^{2r}}{(r+s)!} \frac{(2r+2s)!}{(2r)!} \right]$$

$$= (-1)^s u^{-(2s+1)} \left[D_x^{2s} v^{2s+1} D_v^{2s} \sum_{r=0}^{\infty} \frac{(-1)^{r+s} v^{2r+2s}}{(r+s)!} \right]$$

$$= (-1)^s u^{-(2s+1)} [D_x^{2s} v^{2s+1} D_v^{2s} e^{-v^2}],$$

since $D_v^{2s} v^{2m} = 0$ for m less than s.

Likewise

$$\sum_{r=0}^{\infty} \frac{(2r+2s)!}{(2r+1)!} \frac{(-1)^r u^{2r+1}}{(r+s)!} [D_x^{2s-1} g^{-(2r+2s+1)}]$$
$$= (-1)^s u^{-2s} [D_x^{2s-1} v^{2s} D_v^{2s-1} e^{-v^2}].$$

Therefore, finally,

$$\frac{I_n(u)}{(Y\sqrt{N})^N \sqrt{\pi}} = \cos(u\sqrt{2N}) \sum_{s=0}^{\infty} \left(\frac{1}{2\sqrt{N}}\right)^{2s} [u^{-(2s+1)} K_{2s}](-1)^s$$
$$+ \sin(u\sqrt{2N}) \left(\frac{1}{2\sqrt{N}}\right) \sum_{s=1}^{\infty} \left(\frac{1}{2\sqrt{N}}\right)^{2(s-1)} [u^{-2s} K_{2s-1}](-1)^{s-1},$$

where

$$(2s)! K_{2s} = [D_x^{2s} v^{2s+1} e^{-v^2} H_{2s}(v)]_{x=X},$$
$$(2s-1)! K_{2s-1} = [D_x^{2s-1} v^{2s} e^{-v^2} H_{2s-1}(v)]_{x=X}.$$

Substituting $\sin(x2u\sqrt{N})$ for $\cos(x2u\sqrt{N})$ in the equations defining A_{2m} and then proceeding exactly as above we derive the corresponding expansion for $I_n'(u)$.

V

To obtain the values of K_{2s} and K_{2s-1} note that

$$Xg^2 = 1 - (x-X)/3X + (x-X)^2/4X^2 - (x-X)^3/5X^3 + \cdots$$

gives, for $x = X = 1/\sqrt{2}$,

$$g = \sqrt{2}, \qquad v = (1/\sqrt{2})u,$$
$$dg/dx = -1/3, \qquad dv/dx = (1/6)u,$$
$$d^2g/dx^2 = (4\sqrt{2})/9, \qquad d^2v/dx^2 = -(1/3\sqrt{2})u,$$
$$d^3g/dx^3 = -88/45, \qquad d^3v/dx^3 = (53/90)u,$$
$$d^4g/dx^4 = (824\sqrt{2})/135, \qquad d^4v/dx^4 = -(211\sqrt{2}/135)u,$$
$$d^5g/dx^5 = -28184/567, \qquad d^5v/dx^5 = (79/7)u,$$

etc.

Therefore

$$\sqrt{2}(u^{-1}K_0) = e^{-\frac{1}{2}u^2},$$
$$36\sqrt{2}(u^{-3}K_2) = e^{-\frac{1}{2}u^2}(u^6 - 6u^4 - 9u^2 + 12),$$
$$7776\sqrt{2}(u^{-5}K_4) = e^{-\frac{1}{2}u^2}(u^{12} - 12u^{10} - 183.6u^8 + 1432.8u^6$$
$$+ 2889u^4 - 10368u^2 + 432), \text{ etc.}$$
$$6(u^{-2}K_1) = ue^{-\frac{1}{2}u^2}(u^2 - 3),$$

$$648(u^{-4}K_3) = ue^{-\frac{1}{2}u^2}(u^8 - 9u^6 - 59.4u^4 + 279u^2 + 54),$$

$$233280(u^{-6}K_5) = ue^{-\frac{1}{2}u^2}(u^{14} - 15u^{12} - 414u^{10} + 4494u^8 + 25152.4u^6 - 168723u^4 - 119340u^2 + 304560).$$

References

1. "Théorie Analytique des Probabilités," Laplace, Livre I, Partie II, Chapt. I.
2, 3. Same, Livre II, Chapt. III, Art. 17.
4. "Elements of Physical Biology," Lotka.
5. "Application du Calcul des Probabilités à la Théorie du mouvement Brownien," Hostinsky, *Annales de L'Institut Henri Poincaré,* Vol. III, pp. 47 and 50.
6. "Théorie Analytique des Probabilités," Livre I, Partie II, Chapt. I, Art. 25.
7. "Sur Un Nouveau Développment En Série Des Fonctions," Hermite, *Comptes Rendus*, t LVIII, 1864.
8. "The Theory of Probability: Some Comments on Laplace's Théorie Analytique des Probabilités," Molina—*Bulletin American Mathematical Society*, 1930.
9. *Transactions of the American Mathematical Society*, Vol. 37, 1935, pp. 339–362.
10. "Commentarii Mathematici Helvetici," Vol. I, 1929, pp. 227–254.

The Relation Between Penetration and Decay in Creosoted Southern Pine Poles

By R. H. COLLEY and C. H. AMADON

Poor penetration of the non-durable sapwood is the most important factor in the decay of creosoted southern pine poles. Over 3000 such poles that had been treated with coal tar creosotes of varying types at thirteen creosoting plants in the South have been critically inspected to determine when and where decay started. The poles had been in line from five to twenty-six years under widely diverse climatic conditions in scattered localities east of the Mississippi River. Ninety-five per cent of the failures were poles in which the creosote had penetrated less than 1.8 inches and 60 percent of the sapwood thickness. No failures were found in poles that had been penetrated more than 2.1 inches and 75 per cent of the sapwood thickness. The current Bell System treating specifications require a penetration of 2.5 inches or 85 per cent of the sapwood thickness. The hazard of failure by decay during the ordinary service life of a line is reduced to a practical minimum in poles produced under these specifications.

Introduction

THE creosoted southern pine pole has been justly regarded as a long-lived unit of plant equipment. However, there have been enough instances of failure by internal decay during the first few years in line to focus attention on the poorer poles and to raise questions about the quality and probable length of service of creosoted poles in general. The data presented in this paper were obtained in the course of an investigation to determine how, when, and where decay starts in creosoted southern pine poles in line, and what proportion of the poles are decaying after different periods of service. The results of the study are of particular significance as a basis for engineering the treatment of poles in a satisfactory and economic manner.

General Conclusions About Decay in Poles in Line

In the sections of the lines that were inspected the incidence of decay was definitely correlated with the depth of penetration of the creosote and the per cent of sapwood penetrated.

When all of the 3102 inspected poles of all ages up to 26 years were taken together:

(*a*) There were 62 failures, all of which had penetration less than 2.1 inches and 75 per cent of the sapwood thickness; and the 62 failures were 2.00 per cent of the total poles inspected; and

(*b*) Of these failures 59, or 95.16 per cent, had penetration less than 1.8 inches and 60 per cent of the sapwood thickness.

All the field evidence indicates that the inspected poles, when the sapwood had been well penetrated with creosote, were practically immune to destruction by wood-destroying fungi for a long time. It

is equally clear that if early failures in line and consequent replacement charges are to be reduced to a practical minimum it is essential to inspect the treated poles closely and to eliminate the poorly treated ones before they are shipped to the Telephone Companies.

The Inspected Lines

The selection of the lines to be inspected was based largely on geographical location without prior knowledge of the condition of the poles. An attempt was made to get as wide a distribution as possible. The lines were located in Florida; in the Piedmont section of North Carolina and South Carolina; in the Appalachian foothills and mountains of Tennessee, North Carolina and Virginia; in the Lake States region in Illinois, Wisconsin and Michigan; and in northern New Jersey. Sections of the chosen lines contained from 100 to 200 or more poles that had been set consecutively in one year. Old records, plus identifying marks placed on these poles when they were treated, made it possible to determine the supplier of the poles and the type of creosote used in treatment.

Method of Inspection

External decay is relatively rare in creosoted southern pine poles, so the inspection methods employed were directed particularly at finding internal decay. The latter occurs as a result of infection by water or air-borne spores that probably enter through checks or cracks

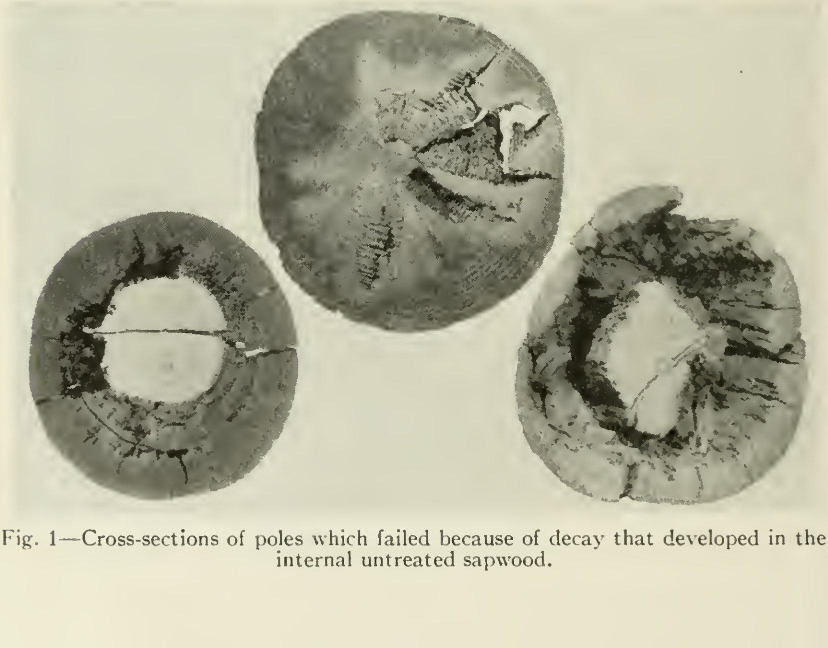


Fig. 1—Cross-sections of poles which failed because of decay that developed in the internal untreated sapwood.

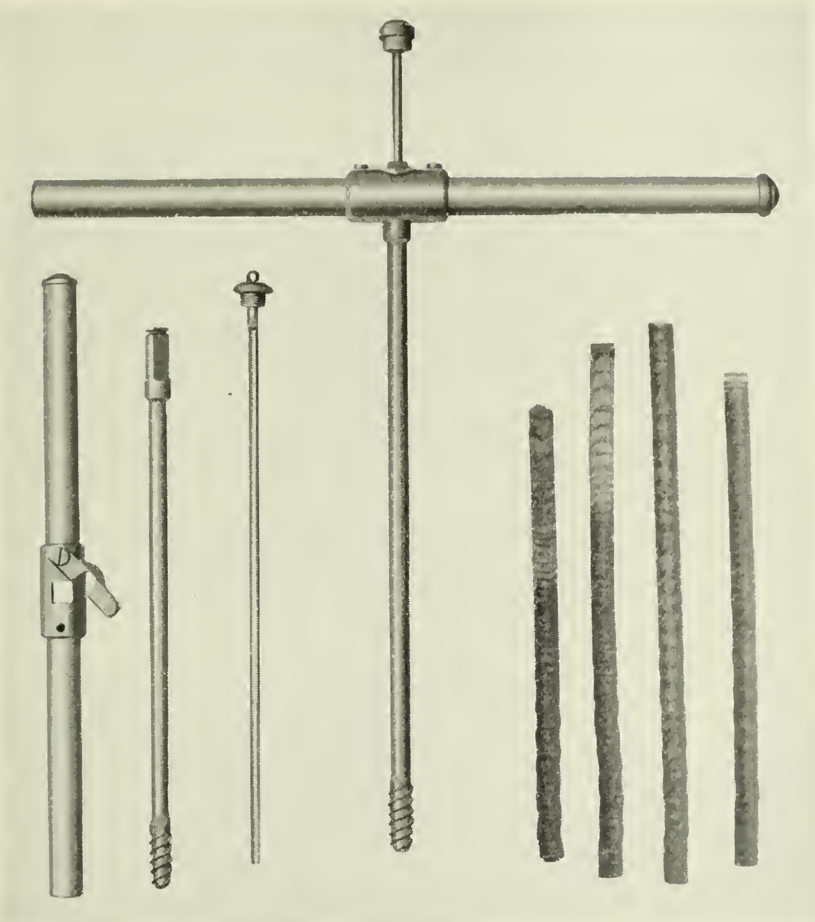


Fig. 2—The increment borer. The central figure shows the borer assembled. At the left are the extractor, the hollow boring tube, and the handle. Four increment borer cores are shown at the right.

and find favorable conditions for growth in untreated, non-durable sapwood lying beneath the treated outer layers of wood. Cross sections of poles showing internal decay of the untreated sapwood are shown in Fig 1.

A systematic inspection of 3102 poles in the selected lines was made by one of the authors, Mr. C. H. Amadon. The possible variances that might arise from different personal methods and interpretations were therefore minimized as far as practicable. Each pole was first tested by sounding with a hammer or a hatchet. When the hammer blows produced a dull lifeless tone, suggesting a hollow or decaying

pole, the suspected region was explored with an increment borer. This tool, which is shown in Fig. 2, is designed to remove a core of wood 0.2 inch in diameter. The increment borer cores were examined to determine the condition of the wood. The depth of penetration of the creosote, the thickness of sapwood, or the thickness of sound shell overlying the decay, were measured on each core if there was any internal decay present. In each decaying pole the shallowest penetration was recorded. Knots, knot holes and woodpecker holes, scars and other surface irregularities or injuries were examined to determine their possible association with the internal decay. If the pole was sound, a single increment borer core was taken at about 4.5 feet from the ground for measurement of penetration and thickness of sapwood.

Results

The results of the investigation are summarized in Tables I and II and in Figs. 3 to 9.

Figs. 3 to 7—Creosoted southern pine poles: Depth of penetration and per cent of sapwood penetrated in relation to decay in twelve-pound full cell poles in line.

Key: Hollow dot indicates decaying pole; solid dot indicates pole failed because of decay; cross indicates heartwood decay or slight external decay below ground line.

PER CENT OF SAPWOOD PENETRATED	DEPTH OF PENETRATION IN TENTHS OF AN INCH																												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25 TO 29	30 TO 39	40 TO 49	50 TO 66
5																												
10			● 1	● 1	● 1																							
15				●● 2	● 2	○ 1	○ 2																					
20				1	● 2		●○● 6		1	● 2		1																
25				1	1	● 2	●○● 3	●●● 5	● 1	3	● 3	1	○ 2															
30							●●● 5	○ 1	●● 6	●●● 6	● 7	● 11		○ 1	○ 1	1												
35							3		1	● 6	●○ 4	5	6	● 2	5	1												
40								1	● 2	○ 5	4	● 12	1	1	5	○ 7	○ 5					1						
45					1		1		●● 3	7	8	○ 4	1	2	11	2	5			2	3	○ 2	2					
50						1				1	○○ 7	○ 5	● 8	3	7	○ 6	● 10			2	5	○ 3	1		1			
55								1		1	5	● 4	5		5	○ 6	6	3	4	6	1	○○ 4	3		2	1		
60										3	1	3			5	6	8	4	9	○ 13	7	6	1		7	3		
65									1	1			5		6	4	12	1	3	7	● 7	6	6	1	12	2		
70										1	1	2		1		2	14	2	5	7	10	13	1		15	7		
75										3					4	4	1	4	8	4	4	9	8		19	9		
80													1		2	3	3	1	5	10	9	12	2		27	16		
85													1			1	1			3	3	6	2	1	19	20	2	
90																	1		1	1	1	1			8	16		
95																										2		
100							2	2	1	5	8	18	19	10	32	+ 54	50	7	47	49	++ 75	107	+ 80	23	++ 322	504	140	18
TOTAL			1	5	7	4	22	10	16	44	48	66	49	20	83	96	116	23	82	104	125	170	106	25	432	579	142	18

Fig. 3—2393 poles in miscellaneous lines less than ten years (average 7.7 years)

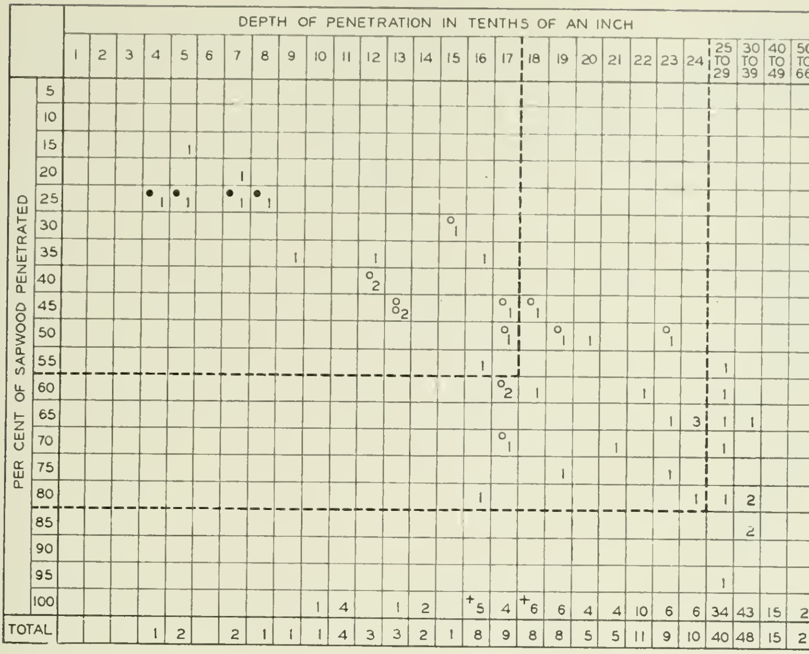


Fig. 4—199 poles in Lynchburg-Savannah line. Age 12 years.

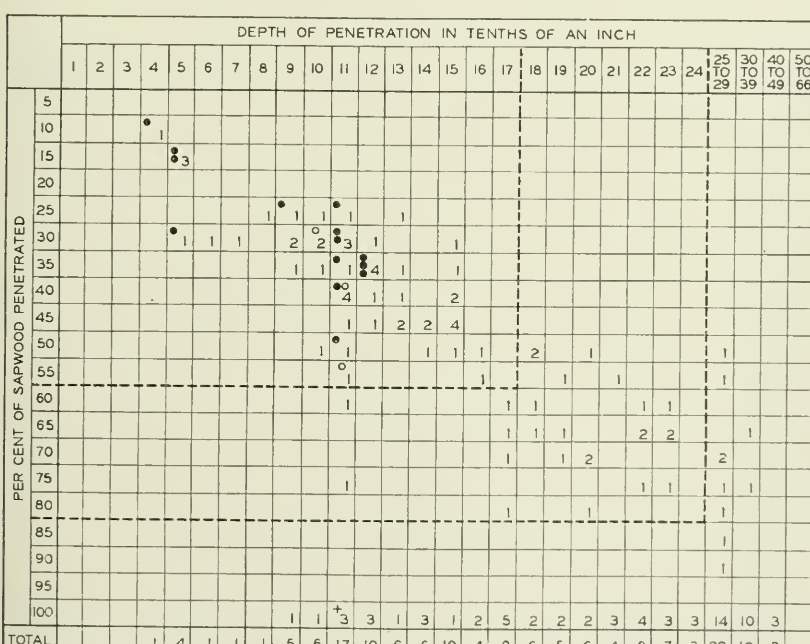


Fig. 5—157 poles in Petersburg-Denmark line. Age 15 years.

PER CENT OF SAPWOOD PENETRATED	DEPTH OF PENETRATION IN TENTHS OF AN INCH																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25 TO 29	30 TO 39	40 TO 49	50 TO 66
5																												
10																												
15						○ 1																						
20									1																			
25																												
30																												
35										1					1													
40																												
45												1																
50										1			1												1			
55																									1			
60															1										1			
65																												
70																1							2		1			
75																	● 1								1	1		
80																	1				1							
85																							1		1			
90																												
95																												
100									1	3	6	8	8	5	5	4	4	5	11	16	16	14	8	4	50	41	6	
TOTAL						1			2	5	6	9	9	5	7	5	6	5	11	16	17	14	11	4	56	42	6	

Fig. 6—237 poles in Jacksonville-Key West line. Age 19 years.

PER CENT OF SAPWOOD PENETRATED	DEPTH OF PENETRATION IN TENTHS OF AN INCH																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25 TO 29	30 TO 39	40 TO 49	50 TO 66
5																												
10																												
15																												
20							○ 1	○ 1			● 2																	
25																												
30															● 1	● 1	● 2											
35									● 1	● 1		1	● 1			1	1	1	2			1	○ 1					
40																				● 1		2		2				
45															1		1					1						
50																		○ 1		○ 1		○ 1	○ 3		2	1		
55																○ 1				○ 1		1		○○ 2	○○ 3	○ 5		
60																	○ 1						1		4			
65																				○ 1						1		
70																										○○ 5		
75																										○ 1		
80																				○ 1					1	3	2	
85																										1	1	
90																										2	1	
95																												
100																			1	1					4	11	12	12
TOTAL							1	1	1	1	2	1	1		2	3	5	2	3	6		6	5	4	14	30	16	12

Fig. 7—116 poles in New York-Scranton line. Age 26 years.

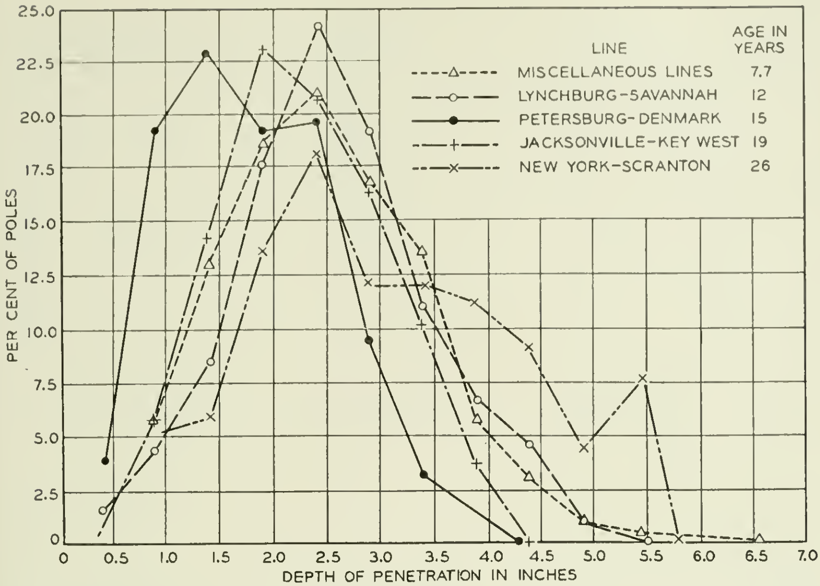


Fig. 8—Frequency curves for depth of penetration in twelve-pound full cell creosoted southern pine poles in line.

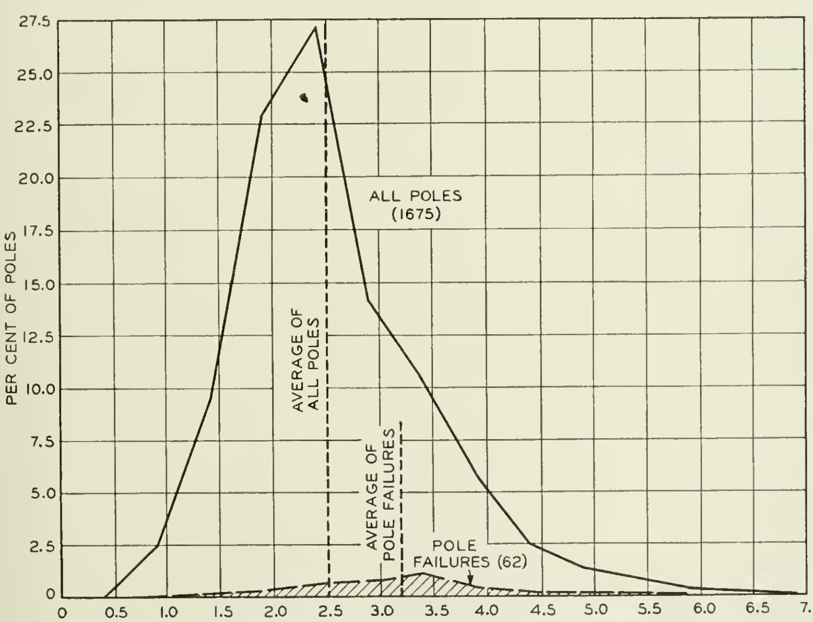


Fig. 9—Frequency curves for sapwood thickness in relation to failure from decay in twelve-pound full cell creosoted southern pine poles penetrated less than 2.5 inches.

TABLE I

Creosoted Southern Pine Poles: Summary of Data on Locality, Age, Creosote, and Penetration, in Relation to Incidence of Decay in 12 Pound Full Cell Poles in Line

	Years in Line	Type of Creosote (Average Residue Above 360° C.)	Number of Poles Inspected	Sound Poles				Poles with Internal Sapwood Decay			
				No.	Per Cent	Average Penetration		No.	Per Cent	Average Penetration	
						Inches	Per Cent of Sapwood			Inches	Per Cent of Sapwood
In Line 10 Years or Less (Average 7.7)											
Florida											
West Palm Beach-Miami Line											
Group 1	8.5	11.14	85	85	100.0	2.4	84.7	—	—	—	—
" 2	8.5	12.79	27	27	100.0	2.2	83.5	—	—	—	—
" 3	8.5	16.57	137	127	92.8	2.1	84.3	10	7.2	1.1	44.5
" 4	8.5	31.12	20	19	95.0	2.4	90.5	1	5.0	0.7	33.0
	8.5		269	258	95.9	2.2	84.8	11	4.1	1.0	43.4
Tennessee											
Nashville-Montgomery Line											
Group 1	5.0	12.00	198	190	96.0	2.3	89.2	8	4.0	1.0	31.0
Nashville-Chattanooga Line											
Group 1	7.0	12.68	24	21	87.5	2.7	77.6	3	12.5	0.8	27.0
" 2	8.0	12.00	151	143	94.7	2.6	88.8	8	5.3	1.3	51.3
" 3	9.0	19.23	25	23	92.0	2.8	88.4	2	8.0	1.7	47.0
" 4	8.0	24.81	112	103	92.0	2.3	86.2	9	8.0	1.3	42.4
" 5	7.0	16.53	57	57	100.0	3.0	97.0	—	—	—	—
" 6	10.0	10.66	106	105	99.0	2.5	95.0	1	1.0	2.2	60.0
	8.3		475	452	95.2	2.7	90.1	23	4.8	1.3	44.6

TABLE I—*Continued*

	Years in Line	Type of Creosote (Average Residue Above 360° C.)	Number of Poles Inspected	Sound Poles				Poles with Internal Sapwood Decay			
				No.	Per Cent	Average Penetration		No.	Per Cent	Average Penetration	
						Inches	Per Cent of Sapwood			Inches	Per Cent of Sapwood
North Carolina											
Greensboro-Selma Line											
Group 1	8.0	31.12	144	143	99.3	3.3	92.7	1	0.7	1.6	43.0
Asheville-Kings Mountain Line											
Group 1	9.0	31.12	51	51	100.0	2.1	73.9	—	—	—	—
" 2	7.0	34.07	23	22	95.7	2.9	94.3	1	4.3	0.7	20.0
	8.0		74	73	98.7	2.3	80.0	1	1.3	0.7	20.0
Asheville-Greenville Line											
Group 1	5.0	34.07	106	106	100.0	2.5	90.2	—	—	—	—
Virginia											
Washington-Lynchburg Line											
Group 1	7.0	24.81	130	130	100.0	2.8	92.8	—	—	—	—
" 2	7.0	19.23	66	62	93.9	2.7	85.4	4	6.1	1.8	41.2
	7.0		196	192	98.0	2.7	90.5	4	2.0	1.8	41.2
Illinois											
Crete-Watseka Line											
Group 1	9.0	12.56	199	188	94.5	2.4	79.1	11	5.5	0.9	33.2
Joliet-Ottawa Line											
Group 1	9.0	11.14	124	122	98.4	2.7	90.4	2	1.6	0.9	27.5
Wyanet-Peoria Line											
Group 1	8.0	11.14	101	101	100.0	2.5	89.5	—	—	—	—

TABLE I—*Continued*

	Years in Line	Type of Creosote (Average Residue Above 360° C.)	Number of Poles Inspected	Sound Poles				Poles with Internal Sapwood Decay			
				No.	Per Cent	Average Penetration		No.	Per Cent	Average Penetration	
						Inches	Per Cent of Sapwood			Inches	Per Cent of Sapwood
WISCONSIN											
Chicago-Minneapolis Line											
Group 1	7.0	12.91	67	66	98.5	2.6	86.1	1	1.5	0.9	42.7
" 2	7.0	12.56	32	32	100.0	2.6	84.4	—	—	—	—
" 3	9.0	31.12	65	59	90.8	2.1	79.9	6	9.2	1.0	29.4
" 4	9.0	10.66	110	109	99.1	2.3	90.6	1	0.9	1.1	34.4
" 5	9.0	16.58	23	23	100.0	2.2	95.5	—	—	—	—
	8.0		297	289	97.3	2.3	87.0	8	2.7	1.0	31.6
MICHIGAN											
Detroit-Ann Arbor Line											
Group 1	9.0	10.66	153	153	100.0	2.4	94.6	—	—	—	—
" 2	7.0	12.91	57	57	100.0	2.2	88.6	—	—	—	—
	8.4		210	210	100.0	2.3	92.9	—	—	—	—
	7.7		2393	2324	97.1	2.4	88.1	69	2.9	1.1	38.4
IN LINE 12 YEARS											
NORTH CAROLINA											
Lynchburg-Savannah Line											
Group 1	12.0	39.76	81	78	96.2	2.8	97.1	3	3.8	1.8	45.5
" 2	12.0	33.68	118	106	89.8	2.6	91.5	12	10.2	1.3	44.9
	12.0		199	184	92.5	2.7	93.8	15	7.5	1.4	45.0

TABLE I—*Continued*

	Years in Line	Type of Creosote (Average Residue Above 360° C.)	Number of Poles Inspected	Sound Poles				Poles with Internal Sapwood Decay			
				No.	Per Cent	Average Penetration		No.	Per Cent	Average Penetration	
						Inches	Per Cent of Sapwood			Inches	Per Cent of Sapwood
IN LINE 15 YEARS											
SOUTH CAROLINA Petersburg-Denmark Line Group 1	15.0	39.76	157	139	88.5	2.0	77.0	18	11.5	1.0	33.0
IN LINE 19 YEARS											
FLORIDA Jacksonville-Key West Line Group 1	19.0	31.59*	237	235	99.1	2.3	96.7	2	0.9	1.1	45.0
IN LINE 26 YEARS											
NEW JERSEY New York-Scranton Line Group 1	26.0	3.30†	116	88	75.8	3.5	77.8	28	24.2	2.0	59.8

* Residue above 315° C.
† Residue above 350° C.

TABLE II

SUMMARY OF PENETRATION AND FAILURE DATA FOR 12 POUND CREOSOTED SOUTHERN PINE POLES IN LINE

Age of Line (Years)	Number of Poles Inspected	Number and Per Cent of Poles Having Penetration Less Than								Per Cent of Poles Having 100% Sapwood Penetration	Failures * Because of Decay in Poles Having Penetration Less Than			
		1.8″ and 60%		2.5″ and 85%		3.0″ and 90%		3.5″ and 90%			1.8″ and 60%		2.5″ and 85%	
		No.	%	No.	%	No.	%	No.	%		No.	% of total poles	No.	% of total poles
7.7	2393	286	11.95	610	24.49	730	30.50	780	32.60	65.73	35	1.46	36	1.50
12.0	199	17	8.54	35	17.59	40	20.10	43	21.61	76.88	4	2.01	4	2.01
15.0	157	55	35.03	81	51.59	88	56.05	90	57.32	42.61	14	8.92	14	8.92
19.0	237	7	2.95	14	5.90	21	8.86	22	9.28	90.71	0	0.00	1	0.42
26.0	116	17	14.66	41	35.95	51	43.95	64	55.19	35.34	6	5.17	7	6.03
Total Poles	3102										59	1.90	62	2.00

* There were no failures in poles having penetration in excess of 2.1 inches and 75% of the sapwood.

Table I is a summary of the data on locality, age, creosote penetration, and incidence of decay for 3102 telephone poles in line in the eastern part of the United States. The pole groups are based on years in service and geographic settings.

Figures 3 to 7 are records of the penetration and the condition of each of the 3102 poles at the time they were inspected. Five age groups are represented, from 7.7 years to 26 years, respectively. These records are graphic illustrations of the fact that the failed poles and the decaying poles had poor penetration. Each solid dot represents a single pole failure, and each hollow dot represents a single infected pole. For example, in Fig. 3 for poles in line ten years or less, the figures and symbols in the 0.8 inch and 25 per cent block mean that there were five poles having penetration 0.8 inch and 25 per cent of the sapwood thickness; and of these five poles two were sound and three were so badly deteriorated that they had to be removed. Similarly in the 0.7 inch and 25 per cent block, two of the three poles were failures and one was infected.

The crosses in the 100 per cent line indicate poles with a little heartwood decay or with slight external decay at the ground line.

The broken lines in Figs. 3 to 7 delimit the individual data for poles having penetration less than 1.8 inches and 60 per cent of the sapwood thickness, and less than 2.5 inches and 85 per cent of the sapwood thickness, respectively. The latter is the minimum penetration called for in current American Telephone and Telegraph Company's specifications for creosoted southern pine poles.

The abscissa in each of the five figures has been warped beyond 2.4 inches in order to condense the charts to reasonable proportions for reproduction. Complete data on the range in penetration for the five age groups are shown in the form of frequency curves in Fig. 8. The 15-year sample from the Petersburg, Virginia–Denmark, South Carolina, line obviously has the poorest penetration of the five groups. It also had the highest per cent of pole failures, as shown by Table II.

Table II is a summary of the number and per cent of the 3102 poles in which the sapwood had been penetrated less than (*a*) 1.8 inches and 60 per cent, (*b*) 2.5 inches and 85 per cent, (*c*) 3 inches and 90 per cent, and (*d*) 3.5 inches and 90 per cent; and it also shows the per cent of poles with 100 per cent sapwood penetration, as well as the penetration in the poles that failed.

Table III contains typical analyses of creosote used in treating the poles.

TABLE III

TYPICAL ANALYSES OF CREOSOTES

Used during the years indicated in treatments of poles represented in Tables 1 to 3

Distillation	1906	1916	1918	1920	1925	1925	1927	1927	1929	1929
0° to 205° C.	12.0	1.86	2.41	3.88	2.98	3.39	2.53	3.47	3.24	1.55
0° to 235° C.	58.0	24.18	4.24	14.33	12.96	24.50	11.75	21.09	14.18	8.14
0° to 315° C.	91.6	68.41	31.86	44.09	43.64	72.08	46.93	66.50	45.95	55.30
0° to 360° C.	96.7*	Not recorded	44.60	59.80	64.77	88.62	67.40	86.85	64.52	82.15
Residue above 360° C.	3.3†	Not recorded	55.10	39.76	34.81	11.14	32.28	12.75	35.10	17.56
Water	Not recorded	Not recorded	1.10	2.7	1.0	1.7	1.9	1.1	1.0	0.7
Tar acids	" "	7.57	1.77	3.0	3.3	7.6	3.1	6.6	2.9	4.3
Sulphonation residue	" "	1.00	2.40	0.5	1.5	2.8	1.8	2.8	1.7	1.0
Benzol insoluble	" "	0.14	1.40	1.9	0.24	0.51	0.42	0.31	0.48	0.29
Specific gravity	1.034	1.052	1.105	1.108	1.079	1.046	1.079	1.045	1.085	1.068

* Distilling to 350° C.
† Residue above 350° C.

Discussion and Interpretation of Results

In reading the following discussion and interpretation of results it should be remembered:

(1) That the poles came from representative areas of the pine forest and from representative treating plants of the South;

(2) That all of the poles inspected were treated in accordance with the process specifications covering a full-cell treatment and calling for a net retention of at least 12 pounds of creosote per cubic foot of wood in the charge; but related studies indicate that the retention in individual poles probably varied from less than 2 to more than 20 pounds per cubic foot;

(3) That the poles were accepted if the treating process and the quantity of oil used conformed to the specifications;

(4) That there was no required inspection at the time the poles were creosoted to determine the results of treatment in terms of penetration and distribution of the creosote in the poles; and

(5) That every pole inspected in the line was the original pole placed in the respective year designated.

The evidence from the field data showed poor penetration to be by far the most important cause of fungous infection and failure by decay. As a matter of fact, the effect, if any, of geographical location or of the type of creosote used was completely masked by the penetration factor.

On account of the wide geographical distribution of lines it was expected that the effect of any definite climatic and meteorological influences on the occurrence of decay would be apparent. It might be expected, for example, that poles set along the warm, moist Florida east coast would be more vulnerable than poles located in the drier north temperate regions. However, the data in Table I for poles in line 10 years or less are not conclusive as to the effect that geographical location may have on the incidence of decay.

The creosotes used conformed as a whole, or in their most important characteristics, to the specifications in effect at the time the poles were treated; but there was a fairly wide divergence in gross chemical characteristics of the oils because of differences in the raw coal tar and in the methods of distilling the tar. Table I includes data on the kinds of creosote used, indicated by the fraction not distilling above 350° C. or 360° C. The data are taken to mean that as far as internal sapwood decay is concerned, the type of coal tar creosote, provided it is a true distillate of coal tar and that it conforms to the specifications, is

less important than the thoroughness with which it is distributed throughout the non-durable sapwood of the poles.

The overall summary in Table I for the 2393 poles in line 10 years or less shows average values for penetration in sound poles and in poles with internal sapwood decay. The summary suggests the existence of the very important relationship between penetration and decay that is definitely shown in detail in Fig. 3. All of the internally decaying poles shown in this figure had penetration less than 2.3 inches and 70 per cent of the sapwood thickness. Furthermore, all except six of these decaying poles and all except one of the poles that failed had penetration less than 1.8 inches and 60 per cent of the sapwood thickness. The group defined by the latter figures may be considered as the "risk group," i.e., poles which by reason of poor penetration may become infected with wood-destroying fungi within 10 years. The 286 poles in this group in Fig. 3 were 11.9 per cent of the 2393 inspected poles that had been in line 10 years or less. The poles making up this 11.9 per cent were possible early failures, but the inspection revealed that only 63 of them, or 22.2 per cent, had actually begun to decay. Of these 63 poles with internal sapwood decay only 35, or 55.5 per cent, failed in service. The distinction between infection and failure is important. In terms of the whole 2393 poles 2.6 per cent were infected with internal decay and only 1.4 per cent failed.

The external decay at the ground line indicated in Figs. 3, 4, and 5 apparently did not exceed one half inch in depth. It was typical of the superficial rot usually found after the ground line of a pole has been raised following a few years of service. During these years the creosote in the exposed outer layers of the wood is depleted, and the favorable moisture conditions at the new ground line facilitate fungous infection of the poorly protected wood.

Another group of poles somewhat above the risk poles in quality may be defined as having penetration more than 1.8 inches and 60 per cent but less than 2.5 inches and 85 per cent of the sapwood thickness. Some of these poles are subject to infection prior to the fifteenth year. The data in Figs. 3, 4, and 5 show that decay developed in 11, or 2.98 per cent, of the poles in this group, and that only 1, or 0.27 per cent, failed in service.

The data in Figs. 6 and 7 and in Table II, show in a striking way the stability of the line when the per cent of poles having penetration greater than 2.5 inches or 85 per cent of the sapwood thickness is relatively large. Not a single pole in this group in the sample from the 19-year old line (Fig. 6) showed any indication of internal sapwood

decay; and in this group in the sample from the 26-year old line (Fig. 7) decay developed in only 6, or 8.1 per cent, of the poles. Moreover, none of these decaying poles at the 26-year age had deteriorated far enough to require removal.

There is no evidence in the data warranting discrimination against thin sapwood poles because of possible extra decay hazard. The average sapwood thickness for all the inspected poles with penetration less than 2.5 inches was 2.52 inches, and the average sapwood thickness for the poles that failed was 3.19 inches. Only 34 per cent, on the average, of the sapwood thickness was penetrated in the poles that failed. When the distribution of the sapwood thicknesses of all the poles with penetration less than 2.5 inches, and the distribution of the sapwood thicknesses of the poles that failed, are plotted as in Fig. 9 there is a clear indication that serious interior decay is more likely to occur in the poorly treated thicker sapwood poles than in the thinner sapwood poles.

The results of the study of actual conditions in line provide a means for evaluating the practical effect of the current specifications for creosoting southern pine poles. The purpose of the specifications is to keep the number of well penetrated poles as high as commercial production will permit, and to eliminate practically all of the poorly penetrated poles at the source of supply. The hazard of failure by decay in poles produced under the specifications appears to be reduced to an economic minimum.

Tandem Operation in the Bell System

By F. M. BRONSON

Tandem operation is becoming of increasing importance in the Bell System. The operating and service features of the different types, and the conditions under which each type is used, are outlined. Charts are included showing, schematically, typical trunking arrangements in the various systems. The increasing use of tandem operation on traffic handled at toll boards is discussed.

THERE are 14,000,000 telephones in the Bell System, served from 6,800 central offices. Means must be provided to permit any one of these telephones to be connected to any of the others. Therefore, facilities must be provided for interconnecting all of the 6,800 central offices. Obviously it would be impracticable to provide direct circuits from each central office to all of the others; this would require $[N \times (N - 1)/2]$, or more than 23 million, groups of two-way circuits, most of which would carry little or no traffic. To keep the number of circuit groups within reasonable limits and to obtain reasonable circuit efficiency, direct circuits are provided only between offices having a sufficient community of interest to justify them. Connections between the others are obtained, as required, through switching operations performed at one or more intermediate points.

The 14,000,000 telephones referred to originate 75,000,000 daily calls, the great bulk of which, of course, are local calls dialed direct by customers or handled at local manual switchboards. There are, however, about 1,500,000 short haul station-to-station toll calls which, because of the close community of interest between the cities involved, are also handled by local operators by methods essentially similar to those used on local calls. Obviously, these are largely concentrated in sections of the country having greatest population densities, such as in the New York City, Boston, and San Francisco metropolitan areas.

To facilitate the interconnection of central offices in areas having large volumes of local and short haul toll traffic, switching arrangements designed particularly for this purpose are frequently provided. These are known as tandem arrangements and, for the purpose of this paper, may be more specifically defined as facilities for the intermediate switching of traffic between central offices other than those facilities involving the use of outward, inward and through toll switchboards and of local switchboards which interconnect trunks of the

ringdown signaling type. More recently, tandem arrangements have been employed in toll offices in connection with the toll lines used for long distance calls, all of which, with a few exceptions, are handled over ringdown signaling circuits.

It is the purpose of this article to describe the operating and service features of the different types of tandem arrangements employed in the Bell System and to indicate the extent to which they are used. Tandem equipment having trunks incoming from other tandem equipment, is a subtandem; some equipments operate both as tandems and subtandems.

A consideration of tandem operation may logically begin with the switching requirements of a single local exchange area. It follows from our definition that a tandem connection involves the cooperation of at least three different offices for its completion. So long as all switching operations are confined within a single office there is, therefore, no occasion for tandem connections. Neither is there any occasion for them when the number and relative locations of the various offices, call them A, B, and C, etc., within the exchange area are such that it is still practicable to handle interoffice calls over direct trunks. With increase in area and number of offices, a point is obviously reached, however, where it is no longer practicable to go, for example, from office A to office U directly, U being located in a remote division of the exchange, although it will still be feasible to go directly between offices A, B, and C, and between offices U, V, and W. Given such an extended exchange area, it will be found to contain some intermediate geographical position at which a tandem office can be profitably located with trunks extending to all the local offices and with switching facilities such that calls from A, B, and C to U, V, and W will be routed to it and will be completed by the interconnection of trunks between the tandem office and these various outlying offices. Such a tandem office would of course be a local tandem office.

Passing from the problem presented by the handling of local traffic within an exchange area, and reserving discussion for later paragraphs, it may be stated as evident that numerous other situations arise within the telephone plant for one or another type of tandem operation. These it is convenient to classify as follows:

I. Manual tandems, at which connections are made manually by plug and jack operation. These include—
   a) Manual straightforward tandems, for completing connections from manual trunks to manual trunks.
   b) Call indicator tandems, for completing connections from dial trunks to manual trunks.

c) Toll office tandems, for completing connections from manual trunks to ringdown toll circuits.
d) Straightforward toll line tandems, for completing connections from straightforward toll circuits to toll switching trunks.
e) Toll switching trunk tandems, for completing connections from manual trunks to manual toll switching trunks.

II. Dial tandems, at which the connections are made wholly by means of switch mechanisms controlled either at the tandem office or at a distant office. These include—
a) Operator tandems for completing connections from manual trunks to dial (or manual) trunks.
b) Full selector tandems, for completing connections from dial trunks to dial (or manual) trunks.
c) Trunk concentrating tandems, for automatically concentrating or collecting traffic which is to be completed over either manual or dial trunks.

Manual trunks include all types of trunks over which the order is passed orally by an operator or by a machine as in the case of call announcer trunks. Dial trunks include those over which the order is transmitted in the form of electrical impulses.

Traffic normally routed over direct straightforward trunks frequently is handled through a tandem system during the night and other hours of light traffic. This is sometimes an economical arrangement since it makes it unnecessary to provide incoming "B" operators during such hours except on positions handling the tandem completing trunks. The speed of connection at such times is substantially as fast as over direct trunks because of the number of "B" positions which it would be necessary otherwise to cover with a small number of operators. Also, a tandem system may be used as an emergency routing during periods when direct trunk groups are out of service because of cable or other failure. Frequently tandem systems are used as overflow routings for traffic normally handled over small direct trunk groups.

Table I indicates the number of the different types of tandem systems in use in the Bell System. In addition to systems of the types shown, tandem operation is obtained through the use of regular local central office equipment in a number of cities where the volume of eligible traffic is very small.

Manual Tandems

Manual Straightforward Tandems

In these tandem systems the incoming and outgoing trunks are of the straightforward type. The incoming trunks are terminated on

TABLE I

TANDEM SYSTEMS AT BELL SYSTEM TOLL CENTERS
(These Constitute Most of the Tandem Systems in Use)

Type of Tandem	No. in Use
Manual Straightforward Tandems	13
Call Indicator Tandems	5
Toll Office Tandems	27
Straightforward Toll Line Tandems	3
Toll Switching Trunk Tandems	2
Panel Sender Tandems—Total	6
—With Operators' Positions	5
Panel Office Selector Tandems	34
Step-By-Step Tandems—Total	61*
—With "B" Board Operators' Positions	5
—With Intermediate Dialing or Key Pulsing Operation	4
Trunk Concentrating Tandems—No. of Cities	9†

* 26 of these have trunks incoming from other tandems.
† With 148 groups of trunk concentrating switches.

single-ended cords on the tandem positions, and, in all but the smallest boards, the trunks are connected automatically to the tandem operator in rotating sequence, a flashing supervisory lamp associated with the trunk indicating to the tandem operator the trunk to which she is connected. When the tandem operator is in this fashion connected to a trunk which an originating operator has selected and over which she wishes to have a call completed, both operators receive momentary tone signals indicating this fact and the originating operator passes the name of the central office desired. The tandem completing trunks appear in the outgoing trunk multiple at the tandem board, usually with idle trunk indicating lamps, and the tandem operator extends the connection from her position by simply plugging the incoming trunk into an idle trunk to the office desired. Plugging into the trunk automatically signals the "B" operator at the called office. The tandem operator's telephone set may be released from the incoming trunk, either by means of a release key provided at her position for that purpose or by the act of plugging into the outgoing trunk. The release key enables the tandem operator to receive a call while establishing the connection on a previous call.

In order to distribute the load and assure the minimum of delay at the tandem board, the various groups of incoming trunks are subgrouped and the subgroups terminated on different tandem positions. In addition, on the larger trunk groups, arrangements are provided so that if the operator upon whose position a subgroup is located is busy, when one or more operators upon whose positions other subgroups terminate are idle, this is indicated to the originating operator in order that she may select a trunk to an idle tandem operator. The number

of trunks handled by the various tandem operators can be varied from hour to hour by means of keys located between each group of 10 cords.

Release of the tandem trunk by the originating operator gives a disconnect signal, simultaneously at both the tandem and "B" boards, and the tandem and "B" operators then take down the connection. The tandem trunk may be reselected for a new call even before the tandem operator has taken down the cord on the previous call.

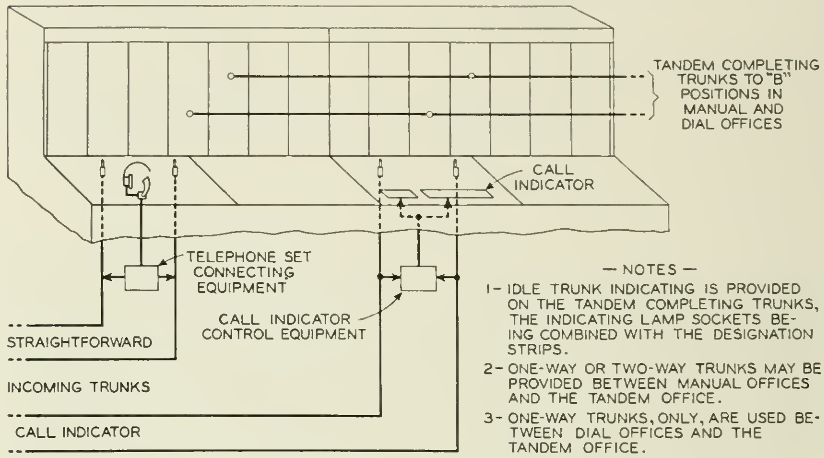


Fig. 1—Manual automatic listening straightforward and call indicator tandem arrangements.

Figure 1 shows, schematically, the circuit and equipment arrangements in the manual straightforward tandem system. Figure 2 is a photograph of the manual straightforward tandem switchboard which serves Detroit, Michigan, and surrounding communities, while Fig. 3 indicates the scope of the Detroit Tandem System.

Where the volume of traffic to be switched is too small to warrant a tandem switchboard, tandem operation frequently is obtained by routing the traffic over straightforward trunks terminating on manual "B" positions at a convenient local office and providing trunks from these positions to other central offices. The operators at these manual "B" positions, therefore, combine the functions of tandem and "B" operators.

Call Indicator Tandems

When manual offices are converted to dial, it is necessary to provide means for completing calls from dial subscribers to all offices, including manual offices, within their local dialing area. The usual arrangement

with respect to manual offices to which direct trunks can be justified, is to display the number dialed by the customer on a call indicator located on a "B" position in the manual office, the operator at this position completing the connection to the called subscriber's line. Where direct trunks cannot be justified, the call indicator may be located on a manual straightforward tandem position, thus forming a "call indicator tandem." The operation at the tandem board is dissimilar in the

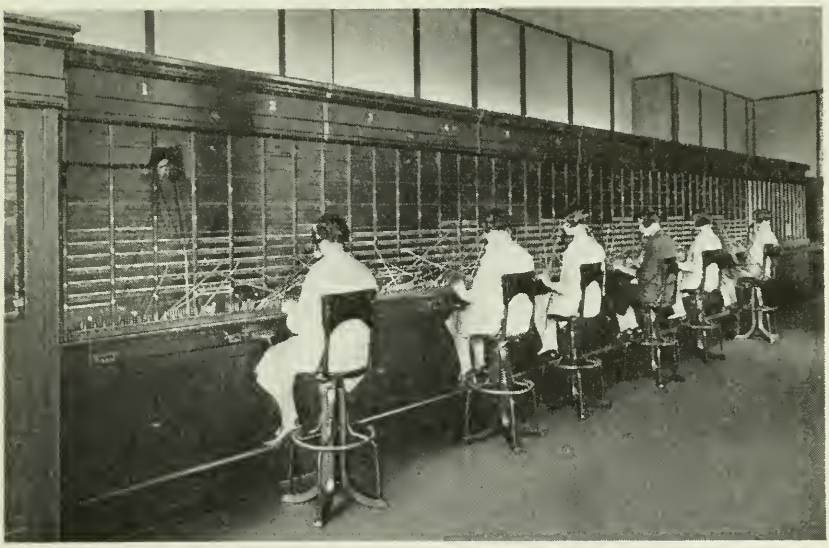


Fig. 2—Detroit manual straightforward tandem switchboard.

following respects to the manual straightforward tandem operation described above: *a*) the display of the central office code and digits of the called number, as dialed by the subscriber, take the place of the name of the central office desired passed orally by an operator; *b*) the tandem operator passes the order orally to the called office; *c*) the tandem operator is not connected automatically to the incoming trunk but, upon receiving a signal on a trunk indicating an incoming call, depresses a display key which connects her telephone circuit to the trunk and causes the number which has been dialed to be displayed on the call indicator.

The call indicator tandem arrangement is shown schematically in Fig. 1, and a photograph of a typical call indicator tandem position is shown in Fig. 4.

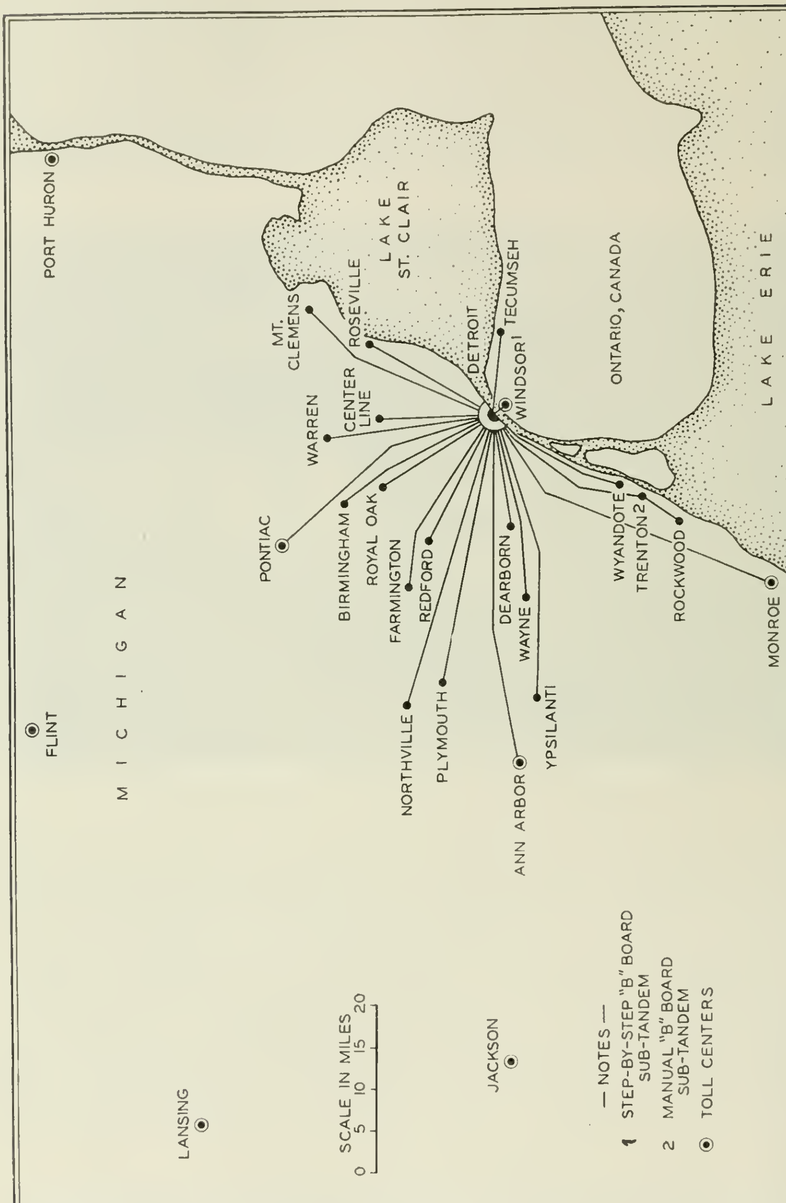


Fig. 3—Detroit tandem area.

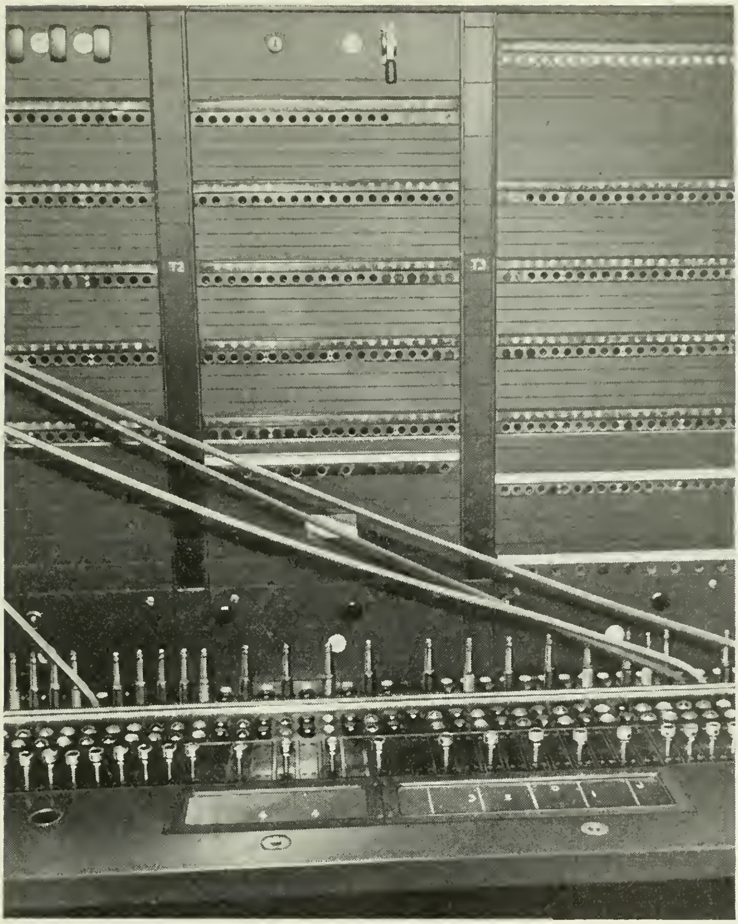


Fig. 4—Call indicator tandem position.

Toll Office Tandems

With the introduction of the combined line and recording method of toll board operation, it was necessary to provide means for giving each outward toll operator access to all of the toll circuits (instead of circuits to certain points only, as required by the former single ticket toll operating method) and in large toll offices where the transmission and switchboard multiple limitations prevent multipling all of the toll lines at the outward positions, toll office tandems provide such a means. Toll office tandems also are used as a means for making toll board circuits available to local operators, at both manual and dial

system "A" boards in multi-operating center cities,[1] to permit the "A" board handling of station-to-station toll calls over such circuits.

Except that they are arranged for establishing connections to ringdown toll lines only and have certain additional features required thereby, the operating and service features of these toll office tandems are similar to those of the manual straightforward tandem previously described. The trunks from the toll board to the tandem positions are usually of the idle-position indicating, idle-trunk indicating, type. Trunks from "A" boards have the idle-trunk indicating feature only.

The ringdown toll lines are multipled in the tandem positions and are equipped with idle-indicating lamps. When a tandem trunk is connected to a toll line, a ring of two seconds' duration is sent automatically, but a ring-release key is provided on each tandem position to permit connection to be made to a toll line without ringing, as is necessary under certain operating conditions. Subsequent rings on the toll line may be made by the originating operator.

When ringdown toll lines appear directly in the multiple before the originating operator and she finds all of the circuits in a particular group momentarily busy, she ascertains when a circuit in the group becomes idle by observing the busy signals associated with the toll line jacks. When connections to the circuits are obtained through toll office tandem equipment, the equivalent of this arrangement for ascertaining when a circuit becomes available may be obtained by providing overflow circuits connected to jacks associated with the different circuit groups in the toll line multiple at the tandem positions. The tandem operator connects the incoming tandem trunk to the overflow circuit and the first toll line in the group to become idle causes a signal indicating this to be given over the trunk to the originating operator. Should the overflow circuit also be busy, the tandem operator connects the incoming trunk to one of a common group of circuits arranged to transmit a signal indicating this condition.

Figure 5 is a photograph of one of the two toll office tandem switchboards in the Long Lines Office in New York City.

In cities not requiring the use of tandem equipment in order to give toll board operators access to the toll lines, a somewhat different toll tandem arrangement is provided for giving "A" operators access to the toll board circuits. Under these conditions the tandem operators' positions are located in line with the toll positions, and the incoming trunks may or may not have the automatic listening feature described in connection with manual straightforward tandems. If not, the

[1] A multi-operating center city is one sufficiently large to require the local operating to be distributed between two or more buildings.

tandem operator connects herself to the trunk and gives the order tone to the originating operator by operating a key associated with the trunk. The ring-release and overflow features are not provided.

Fig. 5—Toll office tandem, No. 1—Long Lines office, New York City.

In a few cases where the volume of traffic to be handled by "A" operators over toll board circuits is very small, a form of toll office tandem operation is obtained, without the use of tandem positions of the types described above, by terminating automatic signaling trunks from the "A" boards on jacks and lamps at the outward toll board positions and having the connections between the trunks and the toll circuits made by means of the regular pairs of cords. The answering jacks are multipled at a number of the toll positions and none of the features normally associated with toll office tandem equipment are provided. The toll board operator answers on the trunk verbally and after receiving from the "A" operator the name of the place desired, establishes a connection to the toll line and rings the distant office. From this point on, the "A" operator handles the call in essentially the same manner as when regular toll tandem equipment is used.

Straightforward Toll Line Tandems

While ringdown operation is the general rule at toll boards, there are a few toll board circuit groups which are operated on a straightforward basis, notably the terminal circuits between New York and Philadel-

phia. These straightforward toll lines are arranged for one-way operation and terminate on single-ended cords at automatic-listening tandem positions in the Long Lines offices in New York and Philadelphia. Regular toll switching trunks are multipled at the tandem positions for reaching the various local offices on incoming calls; idle trunks are found by the tandem operators by tip test, no visual busy signals being provided. The order for connection to the called station is given to the "B" operator at the local office by the originating operator. Ringing on the toll switching trunks is controlled by equipment in the toll line circuits. Switchhook supervision from both called and calling station is received by the originating operator as in local tandem systems. The disconnect signals at the tandem board are controlled by the originating operator. The trunks from the tandem board to the local offices are of the straightforward type. In the case of dial central offices, these trunks terminate on selectors, but an operator at an associated "B" position sets up on a key set the connection to the called subscriber's line.

Toll Switching Trunk Tandems

In a number of the larger toll offices, equipment limitations prevent the multipling of toll switching trunks to all of the local offices at all of the outward positions. Each operator has direct access in the multiple at her position to trunks to the offices from which she normally receives calls. Occasionally, however, it is necessary for operators to reach subscribers connected to other local offices, and to permit this a toll switching trunk tandem is provided. The trunks incoming to the tandem positions are of the cord-ended, key-listening, straightforward type. The originating operator passes to the tandem operator the name of the local office desired.

Dial Tandems

Dial tandems receive calls from operators and, in panel areas, from subscribers also, and are of several types.

Where there is a considerable concentration of short-haul toll traffic within an area served by two or more dial tandem systems which individually serve limited areas, it is sometimes desirable to interconnect such systems and to route calls through one or more of these tandem centers, as required. An example of this is shown in Fig. 9.

Panel Tandems

Panel tandem systems employ panel selectors and are of two general types; one known as the panel sender tandem, and the other as the

office selector tandem. Both are designed for use in cities employing panel type central offices, and therefore these tandems are used to serve only the larger cities and their environs.

Panel sender tandems use senders associated with the tandem equipment to control the electrical operation of the system. The completing trunks may be of the dial, call indicator, or call announcer types. The sender is a device which receives the impulses from the incoming trunk or tandem operators' positions, determines the routing for the call, and sets up the necessary electrical conditions for operating the panel type equipment in the tandem office and the associated equipment in the completing trunks.

These systems ordinarily include both operator tandem and full selector tandem equipment, the latter for traffic routed directly through the selectors from dial system "A" boards equipped with key sets. Local calls dialed directly by subscribers are also routed through full selector tandems when the volume of such calls is so small as not to warrant direct trunks between the originating and terminating offices.

While the equipment arrangements of the full selector tandem are such as to permit operators at dialing type dial system "A" boards (as distinguished from boards equipped with key sets) to dial numbers at the distant offices direct, it usually is more economical to route calls from these boards through the operator tandem positions. Manual "A" boards in panel areas usually are not equipped with either dials or key sets.

All of the trunks to a panel sender tandem office terminate on selectors, but on incoming calls, other than those from operators at switchboard positions equipped with key sets, or dialed by subscribers, the trunk automatically is connected to an idle tandem operator who sets up the called number on a key set provided with a row of keys for each letter and digit in the number. The tandem operator's position then is released automatically from the connection. On connections completed through either the operator tandem or the full selector tandem, disconnection, both at the tandem office and at the called office, is under control of the originating operator, or calling subscriber on calls dialed direct.

The use of call announcer trunks, which at present are designed for panel tandem systems only, permits tandem connections to be completed by key set or dialing operation to small outlying manual offices where, for equipment or other reasons, call indicators are not provided. At the called office call announcer trunks are similar to straightforward trunks, but are so arranged that the connection of the "B" operator's telephone circuit to the trunk causes the call announcer

equipment at the tandem office to reproduce over the trunk by means of a talking film the digits of the number previously set up by the originating or tandem operator.

Trunks may also be provided from panel sender tandem offices to local central offices having step-by-step equipment, to manual tandems, and to dial tandems of either the panel or step-by-step type.

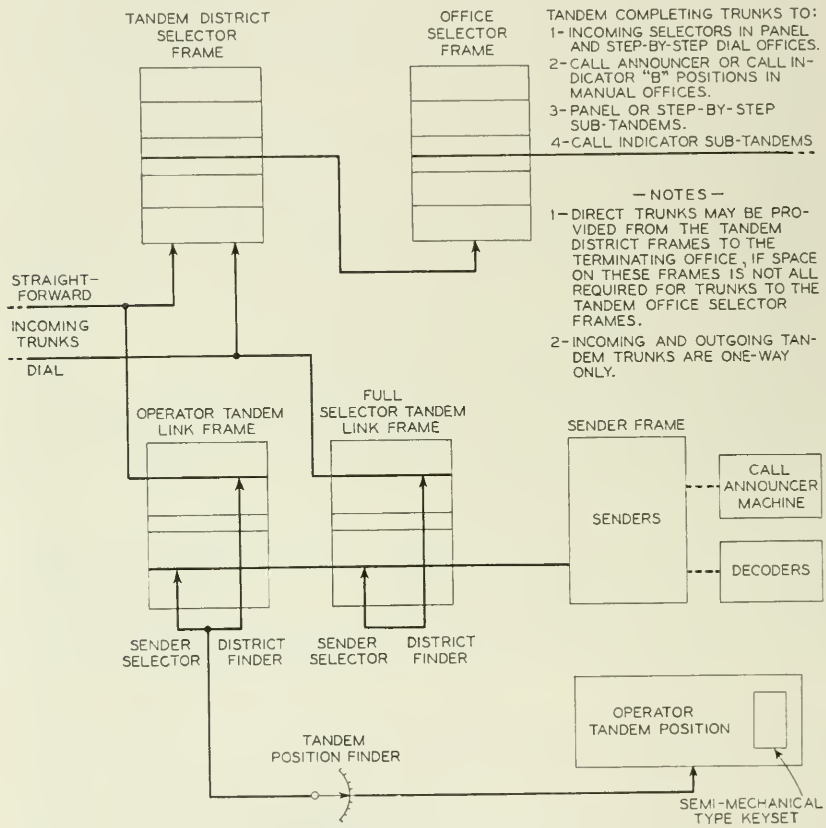


Fig. 6—Panel sender tandem system.

Figure 6 shows schematically the latest type of panel sender tandem system. Figure 7 is a photograph of the operators' positions at Suburban Tandem, one of the panel tandem systems in New York City, while Fig. 8 shows the area served by this system.

An early type of panel sender tandem system known as the semi-mechanical system is in use in New York City. It consists both of an operator tandem and a full selector tandem. The former has opera-

tors' positions equipped with key sets having a row of keys for each of the last four digits and the party line letters in subscribers' numbers, and, in addition, there are coordinate routing keys for selecting trunk groups to the various offices either inside or outside the New York City numbering plan area. In the more recent type of panel sender tandem system, trunk groups to offices outside the numbering plan area

Fig. 7—Panel sender tandem operators' positions—Suburban Tandem—New York City.

are given 3-digit codes which do not conflict with office codes within the numbering plan area, and trunk selection is made by setting up these codes on the key set. The trunks to the operators' positions of the semi-mechanical tandem are on a straightforward key-listening basis and do not have the call distributing feature.

Frequently, it is found desirable in cities served by panel central office equipment to consolidate the traffic, originating in one central office building and destined to a number of central offices in one or more distant buildings, over a single group of trunks to a distant office which serves as a distributing point. This is done by placing panel office selectors at the distant point. Such equipment constitutes an office selector tandem and is arranged to complete connections incoming and outgoing over panel dial trunks only. In all cases the connections are set up by dialing on the part of a subscriber, or by dial or key set operation on the part of an operator. The operation of the office selector tandem equipment, in establishing a connection, differs from that of the full selector sender tandem in that it is controlled by the senders

associated with the local central office equipment at the originating office, or by senders associated with the panel tandem equipment in the case of traffic first routed through the latter.

While panel sender tandem systems ordinarily include both operator tandem and full selector tandem equipment, Knickerbocker Tandem in New York City is entirely of the full selector type, being designed for traffic incoming over dial trunks only.

Step-by-Step Tandems

Step-by-step tandem systems employ step-by-step selectors, and, ordinarily, are of the full selector type without tandem operators' positions. Under these conditions, no senders are required at the tandem office and the pulses which select the central office to which connection is to be made are received over the tandem trunk. Connections may be made through a step-by-step tandem system to both dial and manual offices; to the latter by the use of straightforward, call indicator, or automatic-signaling ringdown tandem completing trunks. Call announcer trunks are not used. Release of the tandem trunk at the originating office automatically releases the tandem selectors and at the same time releases the selectors, or (except in the case of ringdown trunks) gives a disconnect signal, at the called office.

Figures 9 and 10 show, schematically, the step-by-step tandem system in Connecticut, over which most of the toll traffic within the state is handled. Similar systems are in use in Southern California,[1] and on a smaller scale, in other places.

In certain cases the increased trunk efficiency of step-by-step tandem operation is obtained, without the necessity of providing a tandem switching equipment, by locating some of the second selectors of local step-by-step central offices in a distant building serving two or more central offices to which calls are to be distributed. These "distant second selectors" combine all of the traffic to the terminating office over a single group of trunks. In other cases, increased trunk efficiency is obtained through the use of certain levels on the regular step-by-step selectors in a distant dial central office on which to terminate trunks to other central offices.

As stated, step-by-step tandems generally receive the controlling dial pulses over the incoming tandem trunks. When it is desired to complete connections through a step-by-step tandem system from manual offices, this may readily be done if the manual positions are equipped with dials. Where the volume of traffic on which dials could be used is

[1] For a detailed description of the design of the Los Angeles tandem system, see paper by F. D. Wheelock and E. Jacobsen, *Transactions A. I. E. E.*, Vol. 47.

associated with the local central office equipment at the originating office, or by senders associated with the panel tandem equipment in the case of traffic first routed through the latter.

While panel sender tandem systems ordinarily include both operator tandem and full selector tandem equipment, Knickerbocker Tandem in New York City is entirely of the full selector type, being designed for traffic incoming over dial trunks only.

Step-by-Step Tandems

Step-by-step tandem systems employ step-by-step selectors, and, ordinarily, are of the full selector type without tandem operators' positions. Under these conditions, no senders are required at the tandem office and the pulses which select the central office to which connection is to be made are received over the tandem trunk. Connections may be made through a step-by-step tandem system to both dial and manual offices; to the latter by the use of straightforward, call indicator, or automatic-signaling ringdown tandem completing trunks. Call announcer trunks are not used. Release of the tandem trunk at the originating office automatically releases the tandem selectors and at the same time releases the selectors, or (except in the case of ringdown trunks) gives a disconnect signal, at the called office.

Figures 9 and 10 show, schematically, the step-by-step tandem system in Connecticut, over which most of the toll traffic within the state is handled. Similar systems are in use in Southern California,[1] and on a smaller scale, in other places.

In certain cases the increased trunk efficiency of step-by-step tandem operation is obtained, without the necessity of providing a tandem switching equipment, by locating some of the second selectors of local step-by-step central offices in a distant building serving two or more central offices to which calls are to be distributed. These "distant second selectors" combine all of the traffic to the terminating office over a single group of trunks. In other cases, increased trunk efficiency is obtained through the use of certain levels on the regular step-by-step selectors in a distant dial central office on which to terminate trunks to other central offices.

As stated, step-by-step tandems generally receive the controlling dial pulses over the incoming tandem trunks. When it is desired to complete connections through a step-by-step tandem system from manual offices, this may readily be done if the manual positions are equipped with dials. Where the volume of traffic on which dials could be used is

[1] For a detailed description of the design of the Los Angeles tandem system, see paper by F. D. Wheelock and E. Jacobsen, *Transactions A. I. E. E.*, Vol. 47.

Fig. 8—New York suburban tandem area.

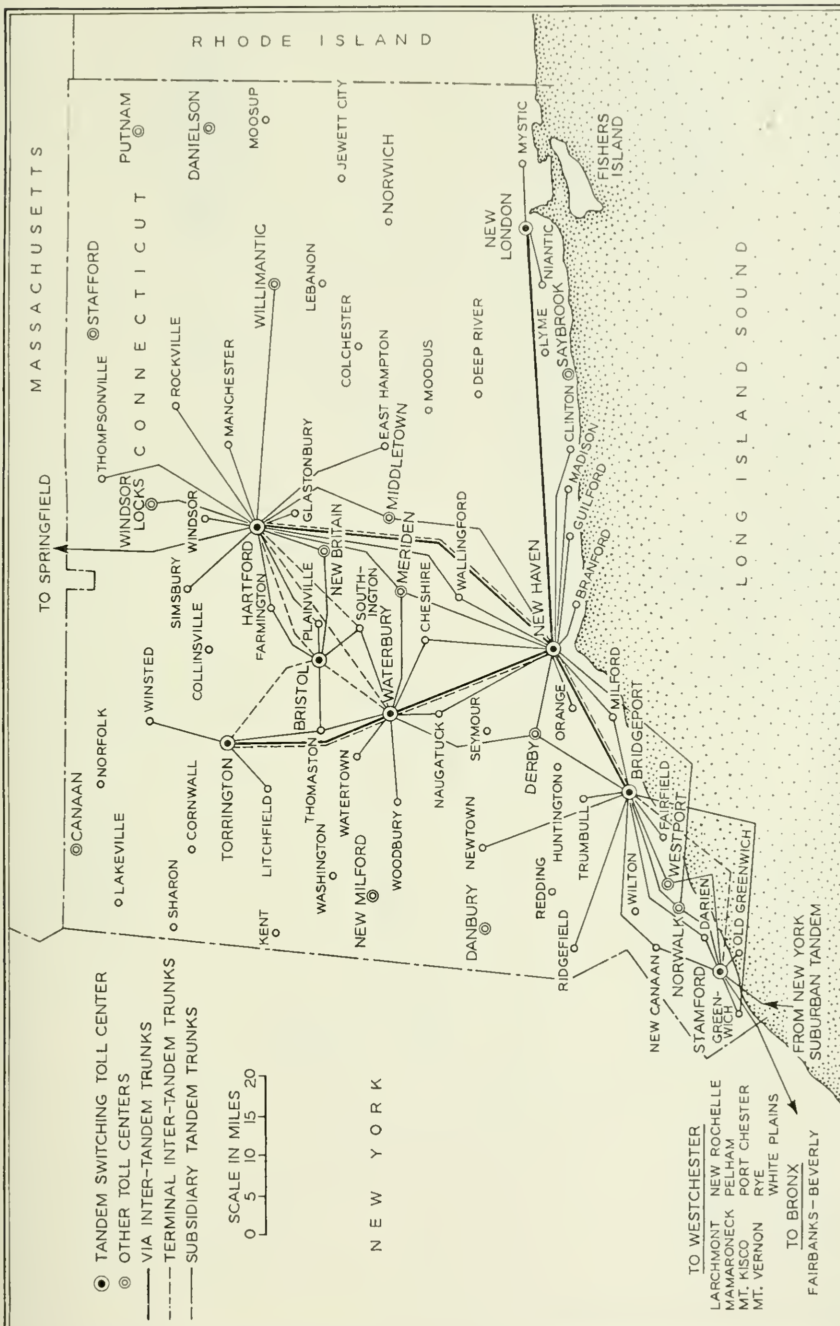


Fig. 9—Connecticut step-by-step tandem area.

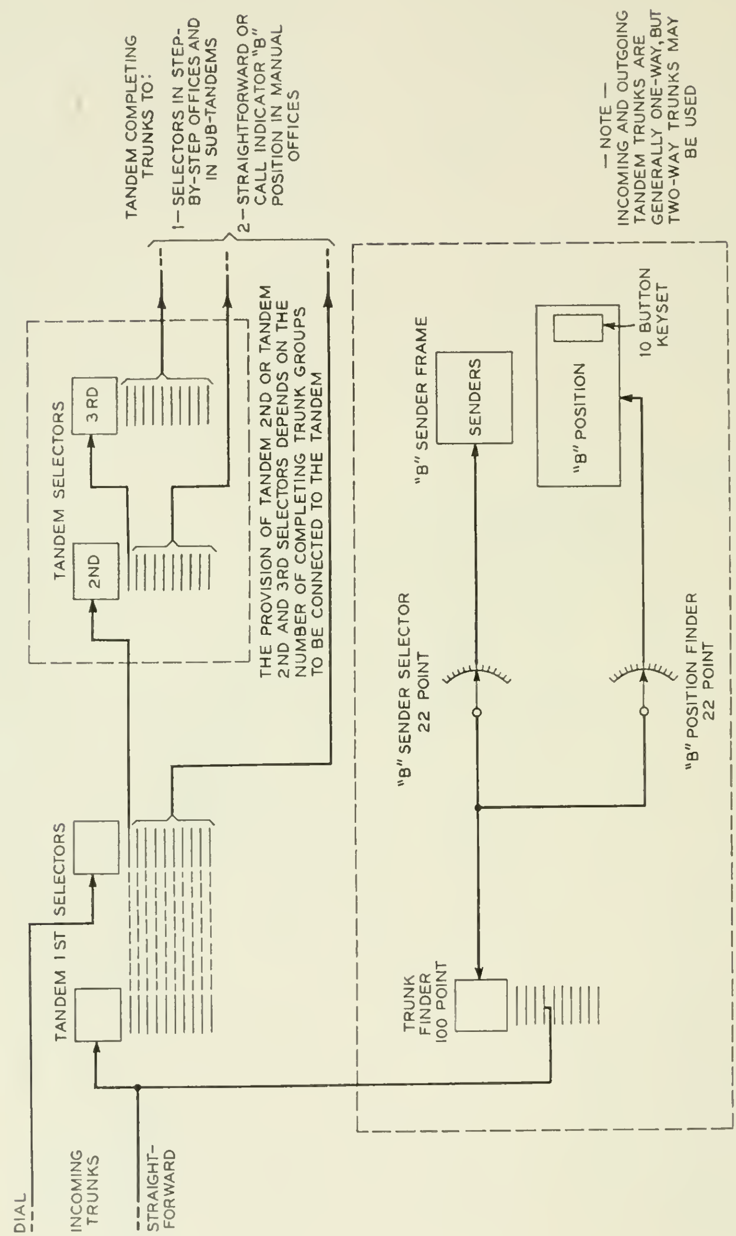


Fig. 11—Step-by-step tandem arrangement with tandem "B" position.

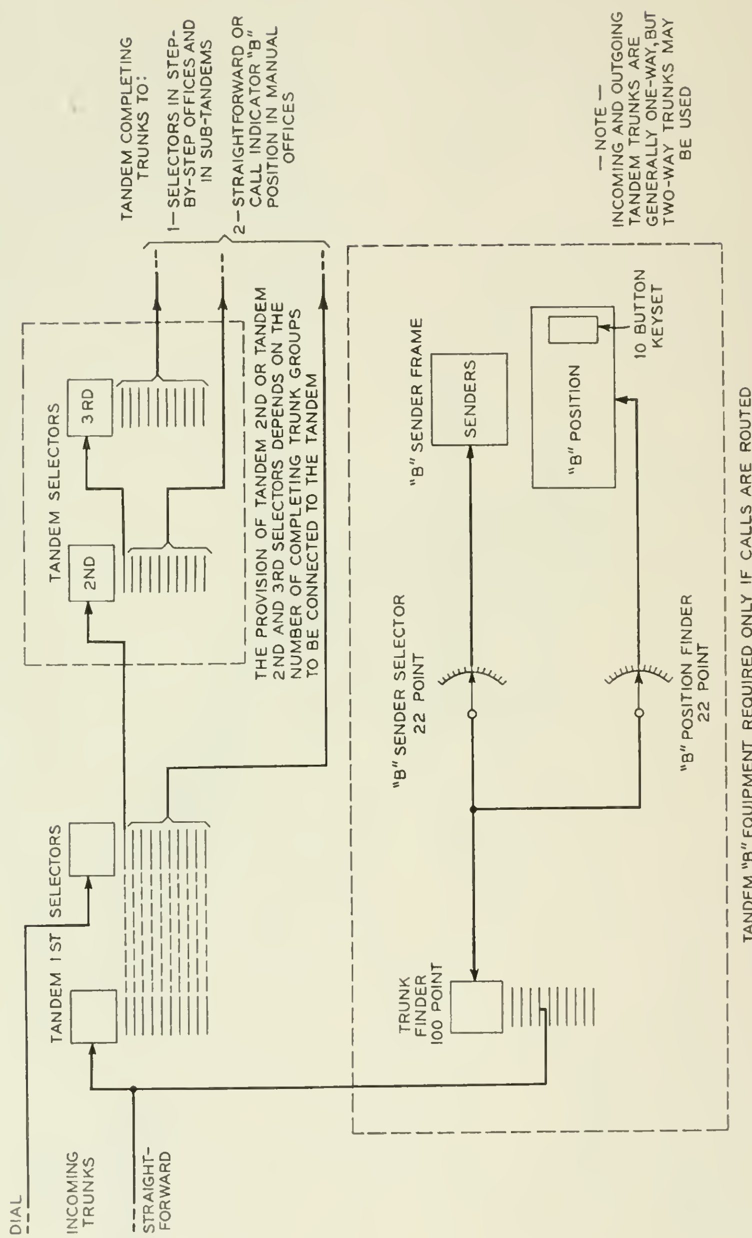


Fig. 11—Step-by-step tandem arrangement with tandem "B" position.

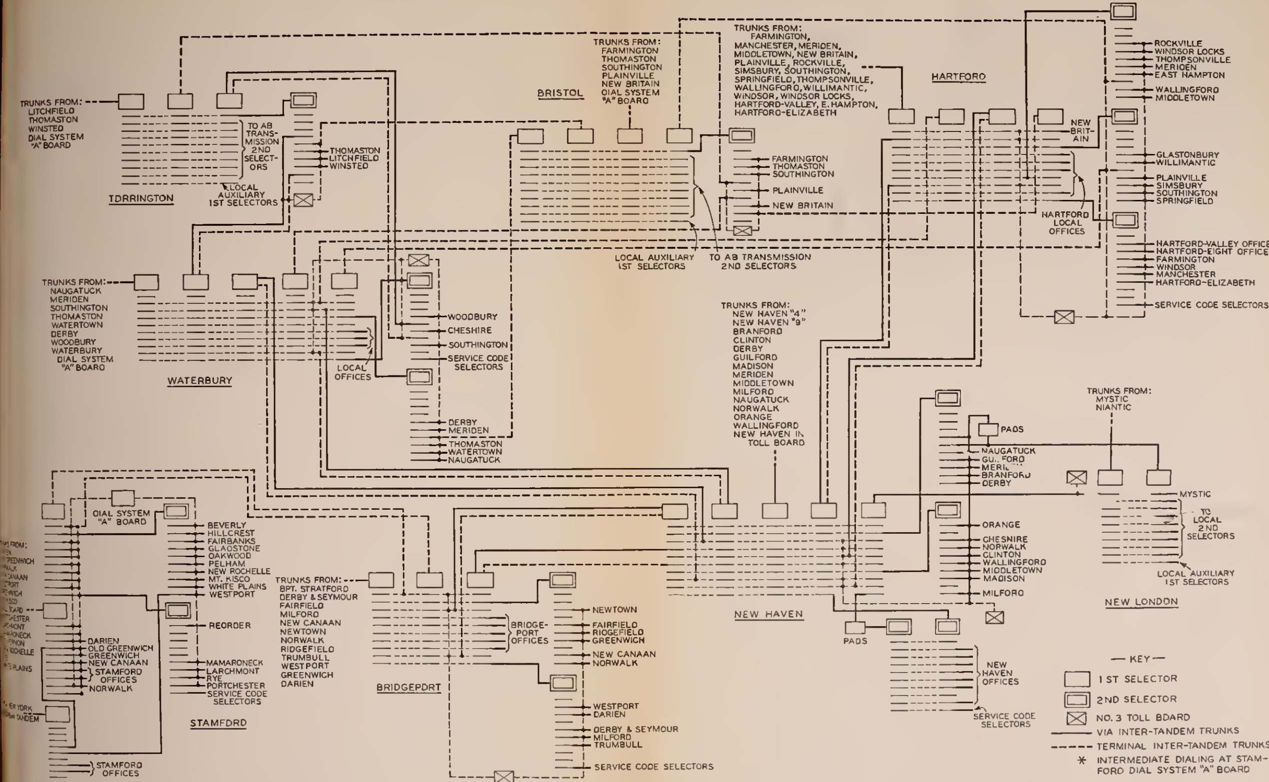


Fig. 10—Connecticut step-by-step tandem system.

relatively small, their provision may not be justified, and other means must be provided for completing connections to dial offices, whether reached over direct or tandem trunks. For tandem operation under these conditions, step-by-step "B" operators' positions, directly associated with the tandem equipment, usually are provided. The incoming trunks are of the straightforward type, and terminate on incoming selectors controlled by means of key sets on the "B" operators' positions and associated senders. An idle "B" operator is connected automatically to the incoming trunk through call distributing equipment and, upon receiving the order and setting up the required digits on her key set, her position is automatically disconnected. The final disconnection of the tandem and local office selectors is under the control of the originating operator. Except for the selectors on which the incoming trunks terminate, all of the selectors are used in common, whether controlled from the "B" positions or by pulses received over dial trunks.

The step-by-step "B" board arrangement sometimes is used as a sub-tandem in connection with manual straightforward tandem operation in a large nearby city, since it provides a convenient means for completing connections to subscribers connected to dial offices within the tandem area.

Figure 11 shows schematically the step-by-step tandem arrangement, including "B" board. Figure 12 illustrates the operators' positions.

Fig. 12—Step-by-step tandem—"B" operators' positions.

It may be mentioned, in passing, that the Connecticut tandem system shown on Fig. 10 does not make use of any step-by-step "B" board equipments.

An arrangement using intermediate dialing or intermediate key pulsing circuits is used in a few cases, in lieu of the "B" board arrangement, as a sub-tandem for completing calls incoming to step-by-step central offices from a manual straightforward tandem system in a nearby large city, such as to dial subscribers in Trenton, New Jersey, from the Newark manual tandem system. Straightforward trunks from the manual tandem switchboard terminate on step-by-step selectors, and on multipled line lamps and answering jacks in the regular switchboard at the incoming end of the trunk, with an auxiliary circuit for lighting the line lamps on an incoming call. When the inward operator plugs into an answering jack in response to a lamp signal, an order tone automatically is sent back over the trunk to the originating operator, who thereupon passes the called number. On key pulsing switchboards, the inward operator sets up on her key set, the desired number, and disconnects from the trunk. On dialing boards, the inward operator dials the called number over a dialing jack associated with the trunk, using a second cord, and disconnects both cords. Release of the connection at the tandem and called offices is under the control of the originating operator.

When used in conjunction with a step-by-step tandem, the intermediate dialing or key pulsing arrangement serves the same purpose, for a limited amount of traffic, as the step-by-step "B" board arrangement described above.

Trunk Concentrating Tandems

Where small volumes of traffic to the same terminating point originate at a number of offices which are closely associated, geographically, trunk costs frequently may be reduced through the use of trunk concentrating switches. Both direct trunks and trunks to a tandem system are treated in this manner.

While different types of switches are used under the various conditions encountered in practice, all function automatically to select a trunk in a common trunk group, or the switches are permanently associated with the common trunks and operate to find the incoming trunk on which a call is waiting. No dial pulses are required to cause the connection between the trunks to be made, and to switchboard operators the outgoing trunks are practically the equivalent of direct trunks to the called office, or to the tandem office, as the case may be.

Figure 13 shows schematically the use of trunk concentrating

switches for giving local operators in Philadelphia access to a common group of trunks terminating on the straightforward toll line tandem in New York City. A more extensive use on intercity traffic is in

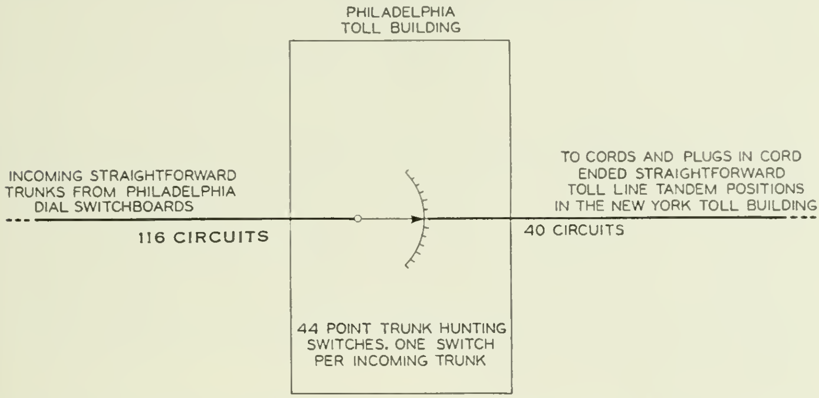


Fig. 13—Trunk hunting switch arrangement for Philadelphia to New York station-to-station "A" board toll traffic.

San Francisco and Oakland, California, chiefly for giving offices in each of these cities direct access to offices in the other city. In San Francisco, 62 incoming trunk groups containing 384 trunks are concentrated on 19 trunk groups and 239 trunks. In Oakland, 188 incoming trunk groups, containing 881 trunks, are concentrated on 25 trunk groups and 333 trunks.

Transmission Arrangements Affecting Traffic Operation

Tandem systems, unlike the general long distance system, operate within definite areas and the associated trunks usually are designed to give satisfactory transmission on connections between any two offices in the tandem area. In some of the larger systems, however, the tandem area is divided into transmission zones and the arrangements provided for traffic between the different zones may require the selection of the proper trunks or paths on the part of the operators.

One arrangement is to introduce telephone repeaters in certain of the paths between the selectors in the tandem equipment, and to route connections requiring repeater gain through these paths. Figure 14 shows schematically the Los Angeles long-haul step-by-step tandem system, in which this arrangement is used. It will be noted, for example, that connections from offices in the Metropolitan Area (Group I) to Long Beach and Santa Monica use no repeaters at the tandem office; and that connections from certain outlying offices

(Group II) are routed through one group of repeaters reached from the "O" level of the incoming 1st selectors, if destined for Long Beach and Santa Monica; and through a different group of repeaters reached from the 6th level of the incoming 1st selectors, if destined for the Norwalk-Artesia-Bellflower exchange area. The transmission gain of this second group of repeaters is higher than that of the first group.

Another arrangement is to use terminal repeaters in the tandem trunks, and to provide pads in certain of the paths between the tandem switches. The longer-haul connections are routed through the tandem switches over paths not containing pads, while terminal and other short-haul connections are routed over paths containing pads. This arrangement is indicated in connection with the New Haven tandem arrangements shown in Fig. 10, where the trunks to New London appear on the 6th level of the tandem second selectors without pads, and on the 7th level with pads.

Still another means of obtaining transmission gain is to provide a second group of trunks to the tandem office over which calls to the more distant offices are routed. For convenience to the operators, these trunks are sometimes designated as a separate tandem system as, for example, Empire Tandem in New York City, which consists of special, high-grade trunks to Suburban Tandem. In the Southern New England System, two groups of trunks are provided between certain of the tandem centers, as shown in Fig. 9, the routing code determining which group shall be used.

Speed of Operation

As might be expected, the speed with which connections can be made through tandem systems varies considerably, depending upon the type of arrangement employed. Table II indicates the relative theoretical speed of operation, in seconds, of some of the more common tandem arrangements. Direct trunks from the tandem equipment to the called office (distant city, in the case of toll office tandems) are assumed; if sub-tandems are involved, a small amount of additional time is required. Also a slight additional time is involved, in the case of step-by-step tandems, if repeaters must be dialed in.

Use of Tandem Systems in Toll Board Operation

While tandem systems have been developed primarily for local and short-haul toll station-to-station traffic handled on manual or dial "A" boards, arrangements have been provided in a number of cities which give toll board operators access to existing tandem systems in

(Group II) are routed through one group of repeaters reached from the "O" level of the incoming 1st selectors, if destined for Long Beach and Santa Monica; and through a different group of repeaters reached from the 6th level of the incoming 1st selectors, if destined for the Norwalk-Artesia-Bellflower exchange area. The transmission gain of this second group of repeaters is higher than that of the first group.

Another arrangement is to use terminal repeaters in the tandem trunks, and to provide pads in certain of the paths between the tandem switches. The longer-haul connections are routed through the tandem switches over paths not containing pads, while terminal and other short-haul connections are routed over paths containing pads. This arrangement is indicated in connection with the New Haven tandem arrangements shown in Fig. 10, where the trunks to New London appear on the 6th level of the tandem second selectors without pads, and on the 7th level with pads.

Still another means of obtaining transmission gain is to provide a second group of trunks to the tandem office over which calls to the more distant offices are routed. For convenience to the operators, these trunks are sometimes designated as a separate tandem system as, for example, Empire Tandem in New York City, which consists of special, high-grade trunks to Suburban Tandem. In the Southern New England System, two groups of trunks are provided between certain of the tandem centers, as shown in Fig. 9, the routing code determining which group shall be used.

Speed of Operation

As might be expected, the speed with which connections can be made through tandem systems varies considerably, depending upon the type of arrangement employed. Table II indicates the relative theoretical speed of operation, in seconds, of some of the more common tandem arrangements. Direct trunks from the tandem equipment to the called office (distant city, in the case of toll office tandems) are assumed; if sub-tandems are involved, a small amount of additional time is required. Also a slight additional time is involved, in the case of step-by-step tandems, if repeaters must be dialed in.

Use of Tandem Systems in Toll Board Operation

While tandem systems have been developed primarily for local and short-haul toll station-to-station traffic handled on manual or dial "A" boards, arrangements have been provided in a number of cities which give toll board operators access to existing tandem systems in

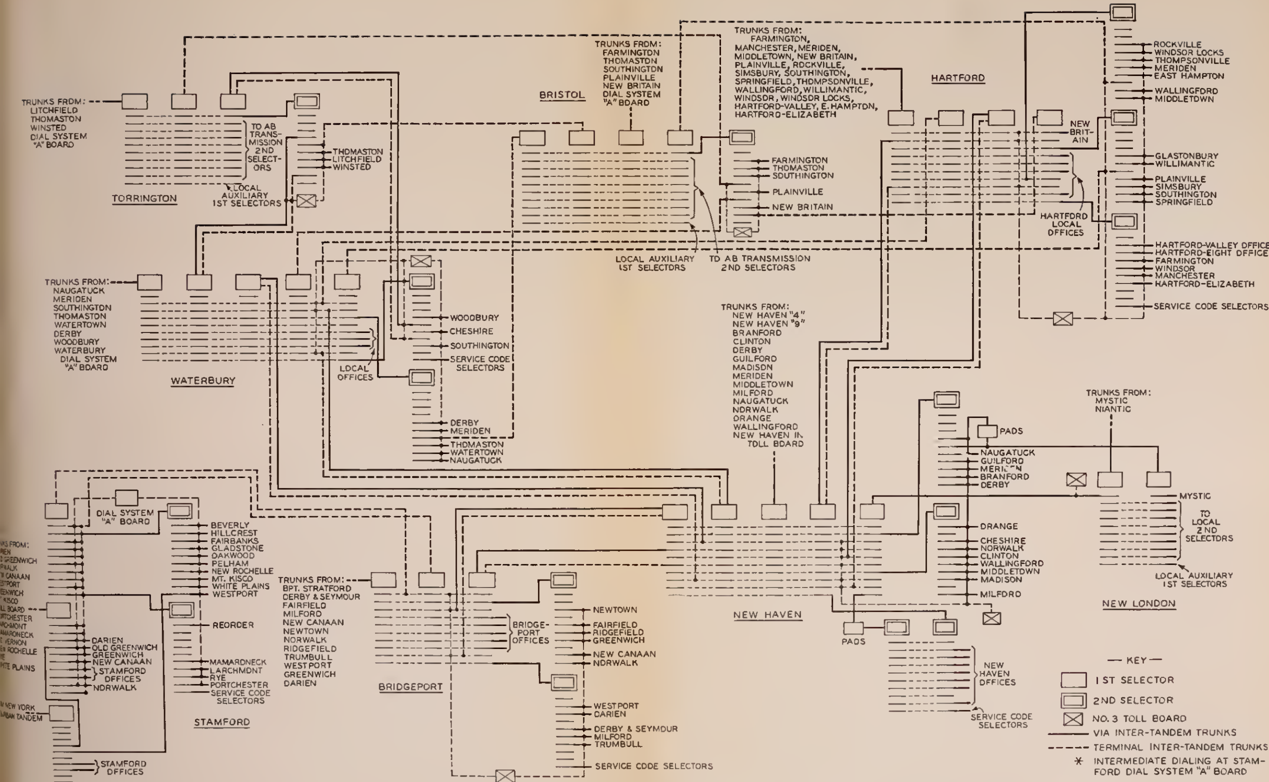


Fig. 10—Connecticut step-by-step tandem system.

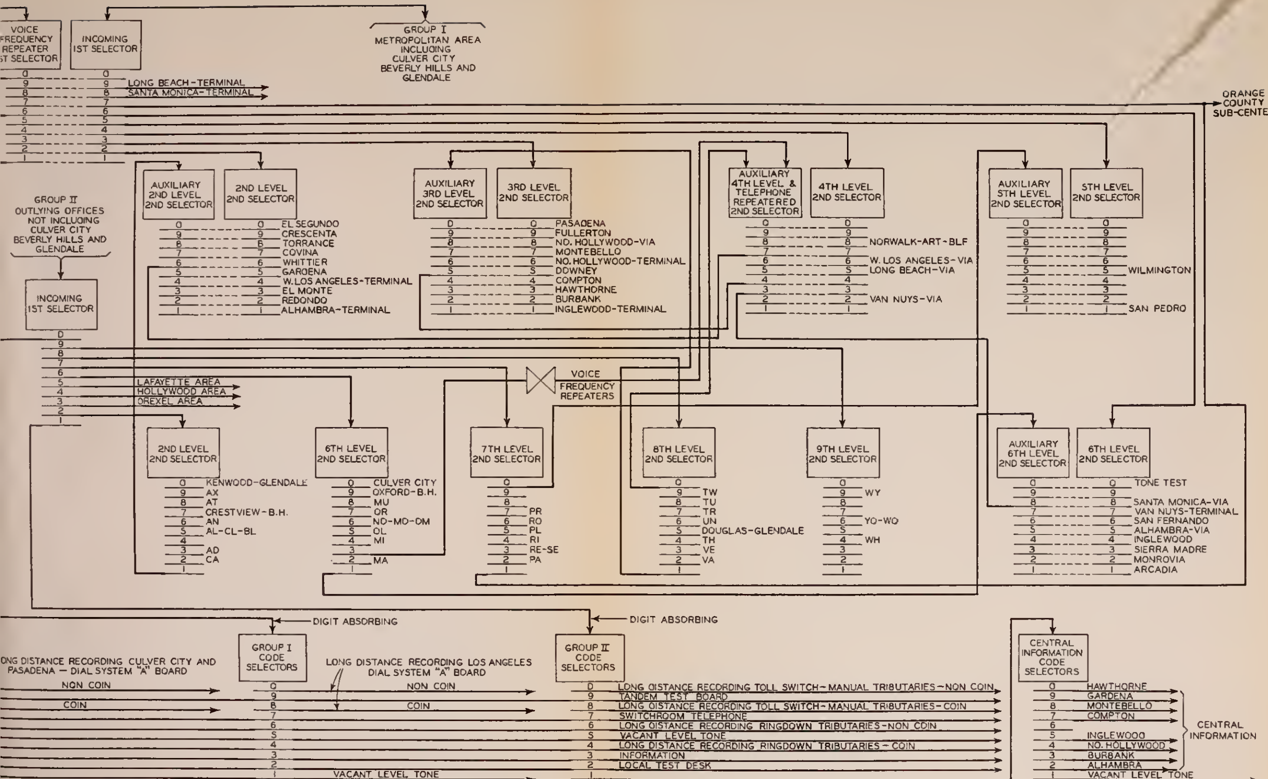


Fig. 14—Los Angeles long-haul dial tandem system.

TABLE II

Relative Theoretical Speed of Operation of Tandem Systems

Type of Tandem	Types of Trunks		Speed in Seconds	
	To Tandem	From Tandem	Calls Handled By Operators*	Calls Dialed Direct By Customers†
Manual Straightforward .	Straightforward	Straightforward—to Manual "B" Pos.	25	
Call Indicator . . .	Call Indicator	Straightforward—to Manual "B" Pos.		31
Toll Office Tandem	Straightforward	Ringdown ‡	29	
Panel Sender Tandem				
Operator Tandem . .	Straightforward	Dial	26	
	Straightforward	Call Indicator	33	
	Straightforward	Call Announcer	34	
Full Selector Tandem				
From Operators at Boards Equipped with 10-button Keysets . . .	Dial	Dial	24	
	Dial	Call Indicator	31	
	Dial	Call Announcer	32	
Dialed by Customers.	Dial	Dial		28
	Dial	Call Indicator		35
Step-by-Step				
Operator Tandem, with Step-by-Step "B" Board.	Straightforward	Dial	22	
	Straightforward	Call Indicator	30	
Full Selector Tandem				
From Operators at Boards Equipped with Dials.	Dial	Dial	16	
	Dial	Straightforward—to Manual "B" Pos.	20	
	Dial	Call Indicator	23	
From Operators at Boards Equipped with Keysets . .	Dial	Dial	17	
	Dial	Straightforward—to Manual "B" Pos.	21	
	Dial	Call Indicator	24	

* On calls handled by operators, this is the interval from receipt of signal at the switchboard to the first ring on the called line. The interval from receiver off hook at the calling station to receipt of signal, is approximately .5 second for manual stations, 3 seconds for step-by-step dial stations, and 7 seconds for panel dial stations. Both of the latter include the dialing by the customer of the "Operator" code.

† In panel areas.

‡ When completed to local multiple in toll board at called office.

order that they may complete person-to-person and other calls to points within the tandem area over tandem trunks rather than over the regular long distance circuits. Tandem systems in general make use of common battery trunks and, since some of the older types of toll switchboards either are not arranged to complete originating calls over such trunks or are not equipped with the dialing arrangements required for dialing through dial tandems, the application of this desirable arrangement is somewhat limited.

As indicated above, the only tandem operation involving ringdown toll board circuits, at present, is for the purpose of making such circuits accessible to operators at switchboard positions not having a multiple appearance of the circuits. It is the belief of the author that, with the further expansion of the long distance plant, the ringdown circuits gradually will be replaced by through supervision circuits—that is, by circuits which, like the trunks used in local and short-haul toll tandem operation, will give the originating operator switchhook supervision from the called station. These new circuits, no doubt, will be arranged for two-way and built-up circuit operation and, incoming to dial areas from toll boards equipped with dials or key sets, will be terminated directly on selectors.

Although changing conditions may suggest better arrangements, it seems probable that, from toll boards not equipped with dials or key sets, the new circuits will be operated on a straightforward basis and terminated on, or controlled at, operators' positions at the incoming end. In the larger cities these positions may have equipment and operating features quite similar to those in the panel sender tandem. In such cities, it may well be that both straightforward and ringdown incoming circuits will be terminated on switches but that the calls will be received, and both terminal and through connections set up, at the operators' positions, thus making use of the tandem type of equipment for all switching purposes in the larger cities. Also, this new inward and through toll office equipment may replace the present type of toll office tandem equipment. In smaller cities, including cities serving manual areas, the incoming circuits may be terminated on equipment having features generally similar to those in the manual straightforward system.

The gradual extension of these arrangements would eventually duplicate, in the long distance toll plant, tandem switching of the type now so extensively used on the local and short-haul toll traffic, and ultimately make the entire United States a super-tandem area.

The scope of the present tandem systems has been determined largely by economic considerations, although the desire to simplify

the service to the customer in the large metropolitan areas also has been an important factor. The general introduction of the tandem type of operation on toll board circuits may affect the economic balance, and except where other factors are controlling, will tend to limit the scope of segregated tandem systems of the present type. It may well be that, eventually, in some of the smaller cities the need for a separate tandem system for the local and short-haul toll traffic will disappear altogether.

Summary

Tandem systems for local and short-haul toll traffic have been provided:

1. To reduce the number of trunk groups required in large metropolitan areas and to insure maximum efficiency on those provided; due consideration being given, of course, to a proper balance between service and costs.
2. To permit the same operating and service arrangements on short-haul toll traffic as on local traffic, thus facilitating the work of the operators and making the service faster, and easier to use by the customer.
3. To reduce the cost of handling large volumes of short-haul toll traffic through the use of toll plant designed to meet the less exacting transmission requirements, as compared with toll plant used for long distance traffic.

These systems vary in type, depending upon the types of local central office equipment which are to be interconnected.

The use of tandem operation in connection with toll board traffic is limited at the present time, but in view of the future volume of this traffic and of new arrangements which now appear feasible, further expansion of the long distance system may be along lines generally similar to those employed in the local and short-haul toll tandem systems; using, of course, operating methods and equipment arrangements which adequately meet the requirements on the longer-haul traffic. These new arrangements may involve both dialing and straightforward toll lines. They may affect the economic balance and, thereby, the scope of tandem systems provided heretofore for local and short-haul toll traffic.

References

In addition to the paper by Messrs. Wheelock and Jacobsen mentioned on Page 10, certain aspects of tandem operation are referred to in the following papers:

1. "General Engineering Problems of the Bell System," H. P. Charlesworth, *Bell Sys. Tech. Jour.*, October, 1925.

2. "Telephone System of the United States," Bancroft Gherardi and F. B. Jewett, *Bell Sys. Tech. Jour.*, January, 1930.
3. "Telephone Toll Plant in the Chicago Region," Burke Smith and G. B. West, *Jour. A. I. E. E.*, January, 1928.
4. "Telephone Trunking Plant in a Metropolitan Area," A. P. Godsho, presented at District Meeting of the American Institute of Electrical Engineers, Philadelphia, Pa. October, 1930.
5. "The Call Announcer," W. H. Matthies, *Bell Laboratories Record*, December, 1929.
6. "Evolution of the Call-Indicator System," E. H. Clark, *Bell Laboratories Record*, December, 1929.
7. "The Manual Tandem Board," C. G. Spencer, *Bell Laboratories Record*, December, 1929.
8. "Toll Tandem Switchboard," C. E. Hokanson, *Bell Laboratories Record*, June, 1930.
9. "Key Display Type Call Indicators," S. T. Curran, *Bell Laboratories Record*, July, 1930.
10. "The Telephone Problem in the World's Largest Metropolitan Area," Kirkland A. Wilson, *Bell Tel. Quart.*, October, 1934.
11. "Dial Switching of Connecticut Toll Calls," W. F. Robb, A. M. Millard, and G. M. McPhee, *Electrical Engineering*, July, 1936.

A Non-Directional Microphone

By R. N. MARSHALL and F. F. ROMANOW

A moving coil microphone is described which responds uniformly over a wide frequency range to sound arriving from any direction. A study of diffraction, the main factor causing directivity of microphones of the pressure type, leads to the conclusion that a small spherical shape is the most desirable for a non-directional microphone. But even fulfilling this requirement in the design of the housing leaves a large directional effect. Hence an acoustic screen has been developed which diminishes diffraction to an extent necessary to make the change in response due to angle of sound incidence imperceptible to the ear. The non-directional microphone is of simple and rugged construction. Adequate precautions have been taken to prevent atmospheric changes from affecting the stability. The small size and unusual shape of the microphone contribute much to its attractive appearance.

IN many situations—such as when a microphone is used as a pick-up for large orchestras or choruses, or in sound picture studios—the sound reaching the microphone directly may be only a small part of the total. Most of the sound arrives at the microphone from directions other than normal to the plane of the diaphragm. If the microphone response differs in these various directions, the output will not truly represent the sound at the point of pick-up—and this is, of course, a form of distortion. This distortion was minimized in the Western Electric 618–A type moving coil microphone [1] by selecting the constants of the instrument so that the field response would be as uniform as possible for sound of random incidence. Still there remained a considerable change of response with the angle of sound incidence and with frequency as is shown in Fig. 2. In the non-directional microphone this variation (Fig. 3) has been greatly reduced so that it is imperceptible to the ear. Moreover, the new microphone is designed to be mounted so that its diaphragm is horizontal.* In this position the instrument is symmetrical with respect to a vertical axis through the center of the diaphragm. If a sound source is placed at some arbitrary location we may rotate the microphone around this vertical axis without changing its response. Hence the instrument is entirely non-directional with respect to the vertical axis. If the microphone is rotated around an axis in the plane of and through the center of the diaphragm a very slight residual directional effect remains and it is this one which has been plotted.

* Since the non-directional microphone is generally mounted with its diaphragm in a horizontal plane the angles of incidence have been labeled 0°, ± 30°, ± 60°, ± 90° retaining 0° as the angle of incidence for sound waves moving in the horizontal plane.

The directivity of pressure type microphones is caused by two factors: (*a*) the variation of the diffraction effect with frequency and with the angle of incidence of the sound wave, and (*b*) the decrease in pressure due to phase shift which occurs when the direction of the progressive

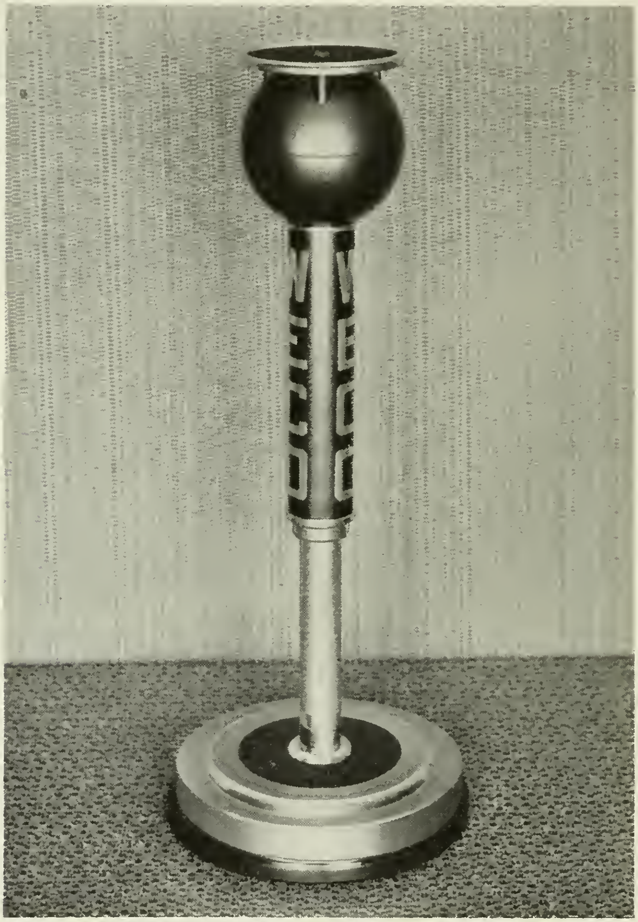


Fig. 1—630-A type moving coil microphone with deskstand.

sound wave has a component parallel to the plane of the diaphragm. Each is also a function of the dimensions of the microphone relative to the wave length of the sound, and in the instruments of the size discussed in this paper the effects become large only at frequencies above 1000 cycles. Directivity might be avoided, therefore, if the microphone could be made small enough; but calculation shows that to make

the effect negligible at 10,000 cycles the instrument would have to be approximately one-half inch in diameter. While a microphone of this size could be built, it is doubtful whether an output level could be obtained which would be adequate for public address, broadcasting, and sound picture use.

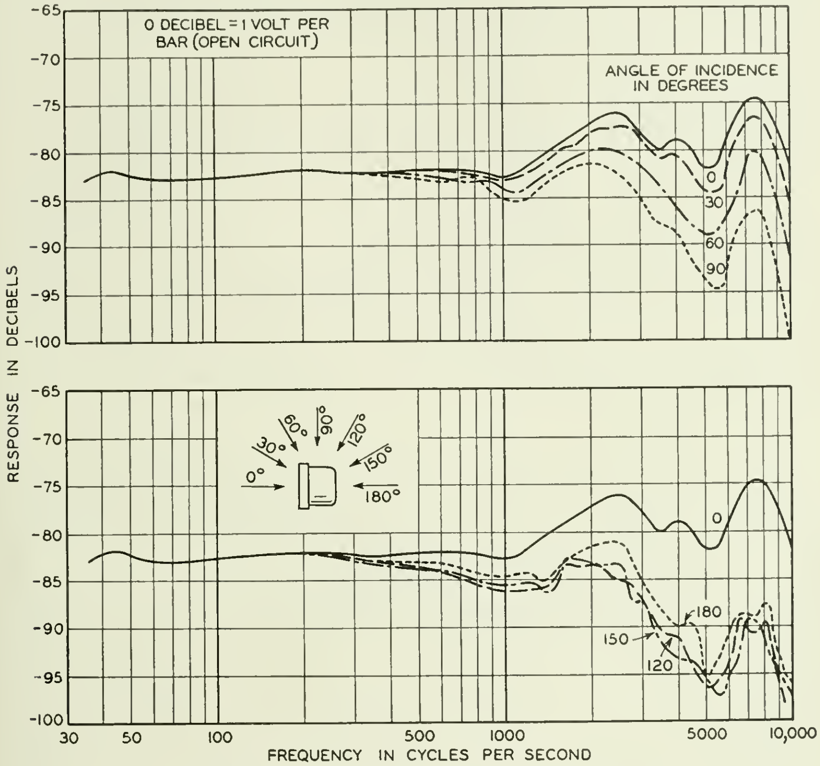


Fig. 2—Field response of the 618-A microphone showing the effect of angle of sound incidence.

Diffraction plays by far the predominant role up to a frequency where the wave-length is comparable to the diameter of the diaphragm, and the effect of phase shift may be neglected. It is principally diffraction which causes the field response of a microphone to differ from the pressure response, and because of the variation of the effect with angle of sound incidence it is impossible to correct for it by adjusting the pressure response. The only alternative is to attack the diffraction problem directly.

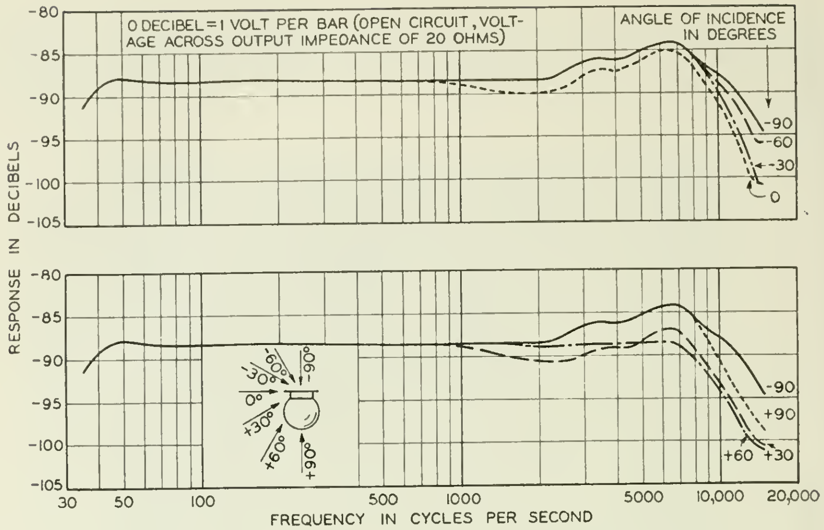


Fig. 3—Field response of a laboratory model of the new 630-A non-directional microphone for several angles of sound incidence.

Diffraction of Sound Around the Microphone

The effect of the shape of the microphone on the directional response has been brought out by study of the diffraction effect of different geometrical objects of equal diameter. The diffraction of a sphere was first treated by Rayleigh [2] and evaluated for a point on the sphere for normal incidence by S. Ballantine,[3] and for other angles of incidence by H. C. Harrison and P. B. Flanders.[4] The effect for a circular plate has been given by L. J. Sivian and H. T. O'Neil.[5] Figure 4 shows the calculated diffraction effect of the cylinder, cube, and sphere * as a function of frequency and angle of incidence in terms of the ratio of the disturbed to the undisturbed sound pressure at a point located centrally in the surface of the object. The abscissae are given as the ratio of the diameter of the object to the wave-length of sound, but the table at the bottom indicates the corresponding frequencies for diameters of 1 inch, 2 inch and 4 inch. If a microphone were built having any one of these shapes, and its diaphragm were made very small and located at the point for which the curves were computed, its response would be increased or decreased approximately in correspondence with these curves as the angle of incidence is changed from + 90° to − 90°. It will be noticed that both cylinder and cube show a marked directional

* The calculated values for the diffraction of the circular plate and cube have been taken from an unpublished work of G. G. Muller of the Bell Telephone Laboratories.

effect, made more serious by the wavy character of the response, while the variation in response for the sphere is much less and the waviness has practically disappeared.

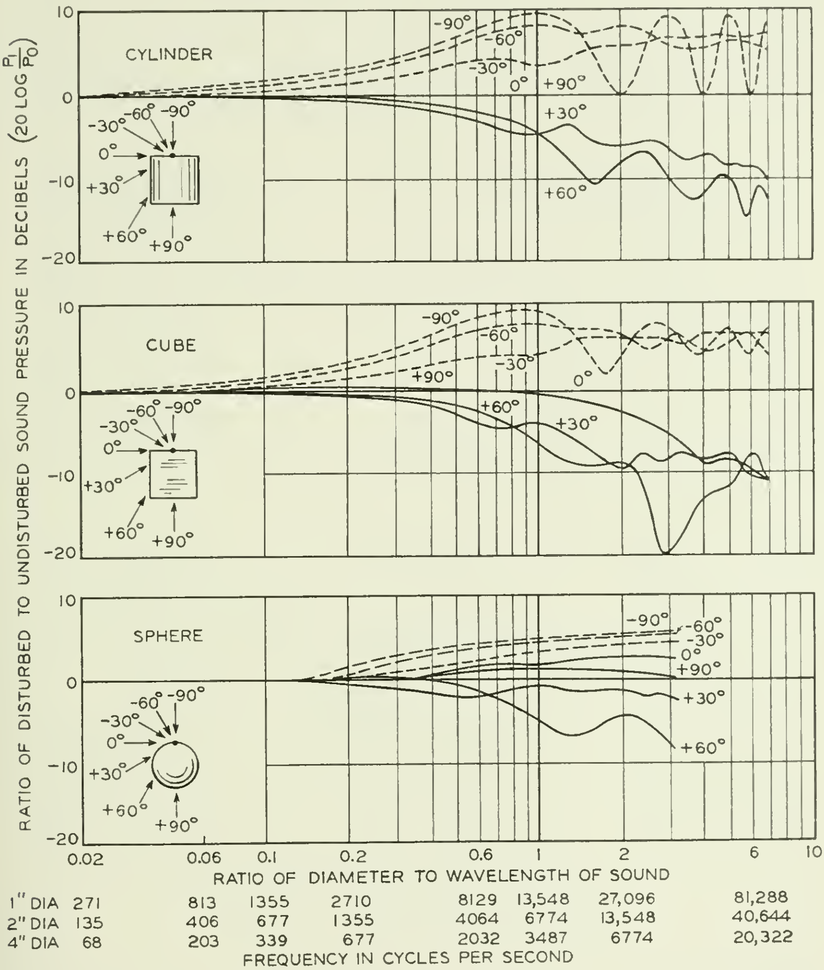


Fig. 4—Calculated diffraction effect at the center of the end form of a cylinder, cube, and sphere for sounds coming from different directions.

It appears then that a very small sphere would be the most desirable shape, but it was found impracticable to reduce the microphone housing below a 2½ inch sphere and to reduce the diaphragm diameter below 1 inch. The field response of the resultant microphone without the acoustic screen for various angles of sound incidence is shown in Fig. 5.

The directional effect, while diminished somewhat compared with that of the 618-A type (Fig. 2), is still not negligible. It was possible, however, to achieve a fairly uniform response with respect to frequency for sound of 0° incidence, that is, sound arriving in the plane of the diaphragm. Since in most instances the direct energy arrives in the horizontal plane, uniform, non-directional response for this important plane may be secured by mounting the microphone with the diaphragm in a horizontal position. Still there is a tendency for the response to be

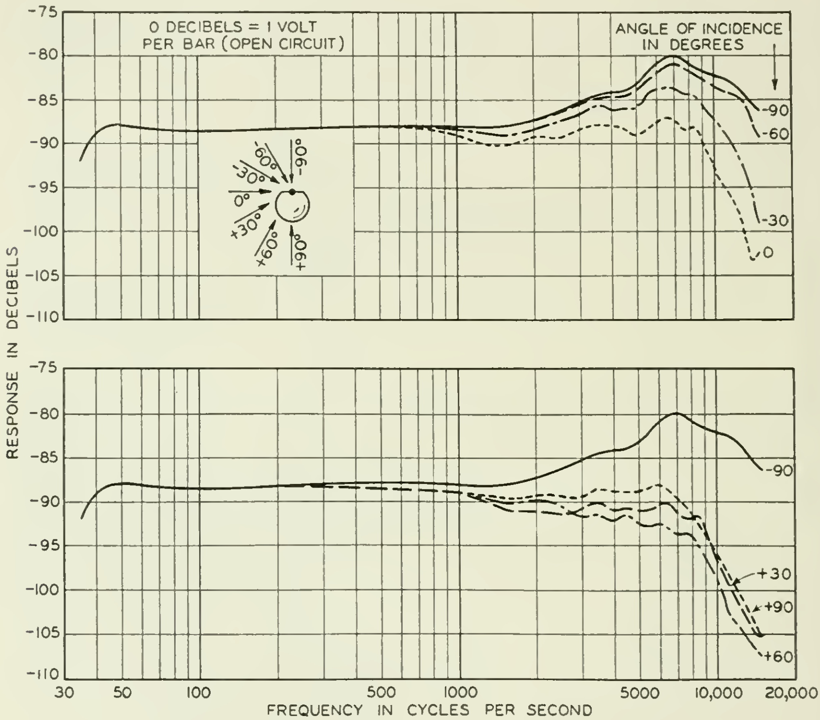


Fig. 5—Field response of a laboratory model of the 630-A non-directional microphone without screen.

too high for high-frequency sounds coming down from above, that is, directly toward the diaphragm, and too low for similar frequencies coming from angles very much below the horizontal.

To determine to what extent diffraction contributes to this residual directivity in the vertical plane let us consider the diffraction of a geometrical shape resembling that of the microphone, namely, a two and one-half inch sphere with a flat face one and one-eighth inches in diameter. We may approximate the diffraction effect for this shape by

combining the known effects for a sphere and flat plate shown in Fig. 6.* Although the sphere is twice the diameter of the circular plate, it is seen that it has the larger effect only at the lower frequencies. The arrows indicate the probable effect to be taken for that of the flat-faced sphere. Although this result applies strictly only at a point at the center, measurements [5] have shown that up to 15,000 cycles for a one and one-

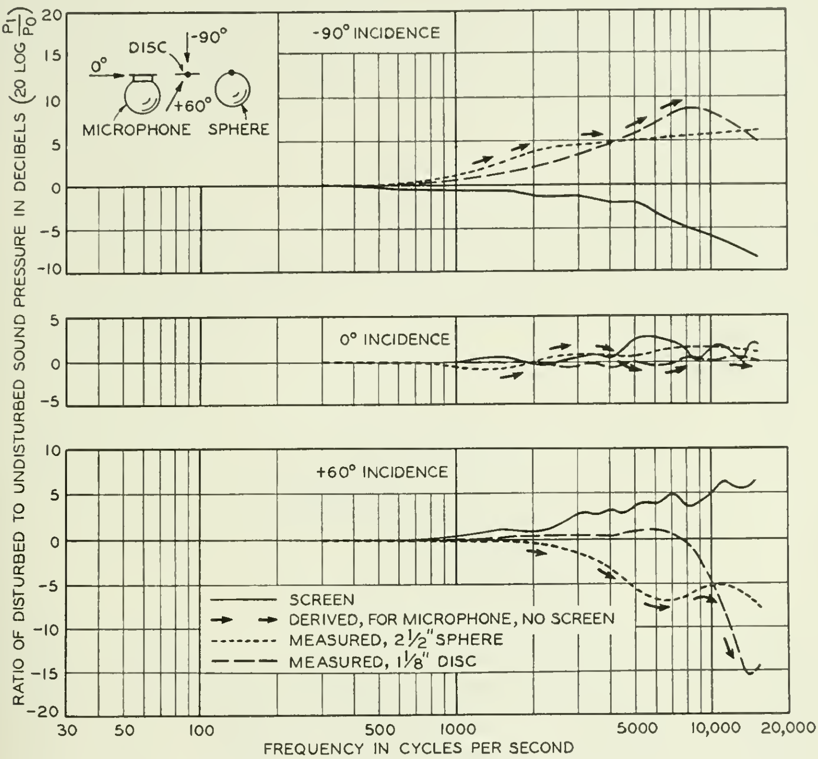


Fig. 6—Measured diffraction effect of a 2½″ sphere, 1⅛″ circular plate, and acoustic screen and derivation of diffraction effect of 630-A microphone without screen.

eighth inch circular disc there is little variation over an area comparable to the effective area of the microphone diaphragm. Hence, the diffraction effect derived in this manner is added to the computed contour pressure response (see Appendix A) to obtain the theoretical field

* The effect at + 60° incidence has been shown as more significant since the diffraction for + 90° is small. The latter effect corresponds to the optical bright spot at the center of the disc on the side away from the light source. This effect occurs over such a small area for angles very close to + 90° that it is of no practical use in this case. However, this does account for the + 90° response of microphones often being higher than the + 60° or + 30° response.

response shown in Fig. 7 where it is compared with the experimental response. The largest deviation between theoretical and experimental values is in the zero degree response at frequencies from 10,000 to 15,000 cycles. This difference is attributable to the factor mentioned earlier, namely, the decrease in effective pressure due to the phase shift of a plane sound wave traveling across the face of the diaphragm. It is of the same order of magnitude as that calculated by H. C. Harrison and P. B. Flanders [4] for a stretched circular membrane. It is concluded, therefore, that the diffraction effect indicated by arrows in Fig. 6 is representative of that of the actual microphone.

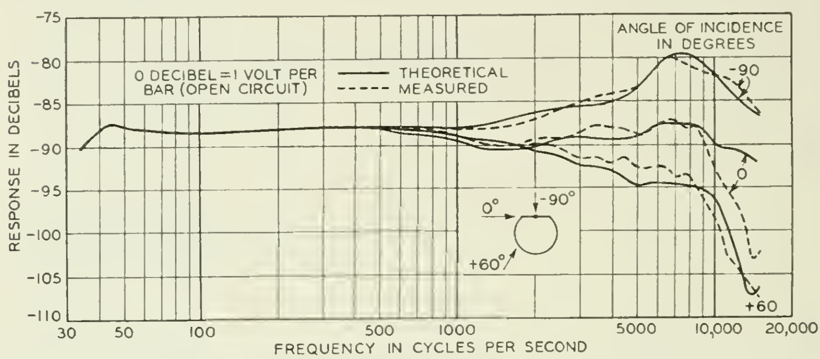


Fig. 7—Comparison of theoretical and experimental field response of a laboratory model of the No. 630-A non-directional microphone without screen.

Effect of Acoustic Screen

From the quantitative considerations of the diffraction of the spherical moving coil microphone without screen it becomes clear that if the microphone is to be made non-directional in the vertical plane also, an element must be introduced which compensates for the increase and decrease in field response due to diffraction.

The screen which was developed for this purpose is a disc 2½ inches in diameter and made of material having a very high resistance-to-mass ratio. This disc is supported approximately 1/8 inch in front of the microphone grid. The diffraction effect of this screen has been measured in terms of the effect on the face of the microphone. Figure 6 gives the effects for sound of 0°, − 90°, and + 60° incidence and compares them with those of the microphone without the screen. From these data it may be seen that the acoustic screen compensates for the microphone diffraction effect, for (1) it has least effect for sound of 0° incidence; (2) it causes a decrease in the − 90° field response; and (3) it causes an increase in the + 60° response.

The variables in a screen of this type are its diameter, impedance, and distance from the microphone grid. It is a combination diffraction and impedance screen; for part of the sound is attenuated by passing directly through the screen while the rest is diffracted. The proportion between the two is a function of the impedance of the screen and of the ratio of its diameter to the wave-length of the sound. At lower frequencies most of the sound coming from the top is bent around the screen while at higher frequencies more of it travels directly through and becomes attenuated. For sound coming from the side, the screen has little effect. When sound comes from the bottom some of it is reflected onto the face of the microphone. The acoustic screen thus makes the instrument non-directional in its response characteristics.

General Design

Besides these changes designed primarily to reduce the directional effects, extensive changes were made in the internal construction and arrangement of the microphone to make the response more uniform and to extend the frequency range. The general construction is shown in Fig. 8. The desirability of making the diaphragm as small as possible

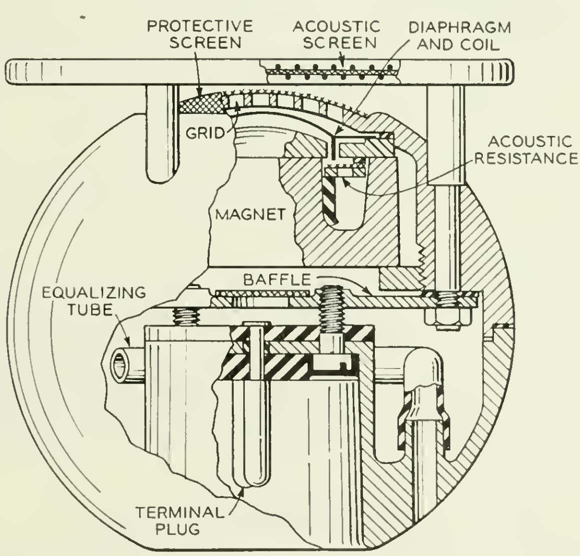


Fig. 8—Simplified cross-sectional view of the non-directional microphone.

has been pointed out in the discussion of microphone diffraction. However, decreasing the size rapidly reduces the sensitivity which is proportional to the area of the diaphragm and the flux density of the

magnetic field around the coil. In order to obtain in this instrument a signal-to-noise ratio sufficiently high for all practical purposes, it was not considered advisable to use a diaphragm smaller than 1 inch in diameter. The loss in sensitivity resulting from this choice was partly offset by making the diaphragm light in weight and of very low stiffness. It is also very important that this diaphragm vibrate as a simple piston throughout the entire range; but to obtain this action over a wide range of frequencies has proved in the past to be a very difficult problem. For this new microphone, a diaphragm was developed which has a rigid spherical center and a tangentially corrugated annulus and which has in addition a high area to stiffness ratio. No evidence of vibrating in other modes is shown by this structure below 15,000 cycles. The diaphragm is cemented to a raised annulus on the outer pole-piece. The outer and inner pole pieces are of soft iron and are welded directly to the magnet which is made of high grade magnet steel. The diaphragm is damped by an acoustic resistance which is supported below the coil by a brass ring. This ring is held in place with rubber gaskets.

The size and shape of the housing were selected with particular reference to the requirements that had to be met. The size is such that the housing fits closely over the diaphragm and thus produces little more diffractive effect than would the diaphragm itself, and the spherical form allows sufficient amount of air space behind the diaphragm, which is essential to minimize the impedance to vibration. To prevent resonance within the case an acoustic resistance baffle is provided to divide the space into two parts. A tube with its outlet at the back of the housing serves the double purpose of equalizing the inside and atmospheric pressures and of increasing the response of the instrument at low frequencies.

In the non-directional microphone the resonance in the cavity in front of the diaphragm is controlled by the design of the protective grid. Instead of being the source of an undesirable distortion, the grid and cavity have become a valuable aid in improving the response of the instrument at frequencies from 8,000 to 15,000 cycles. This grid also incorporates a screen which prevents dust and magnetic particles from collecting on the diaphragm.

Method of Measuring Field Response

The method of making the frequency-response measurements is similar in general details to the method outlined in a paper by W. C. Jones and L. W. Giles [1]. Figure 9 shows the arrangement of the room and testing apparatus. A very small, specially developed, condenser microphone was used in determining the sound field pressure. The

determination of the field calibration of this reference microphone presented an interesting problem and an original solution is described in Appendix B.

At low frequencies where the wave-length of sound is very large compared to the dimensions of the microphone, the coupler shown in Figure 9 is used. At the higher frequencies a steady state sound field is set up in the damped test room and the pressure at a given point is measured by means of the small reference condenser microphone; then

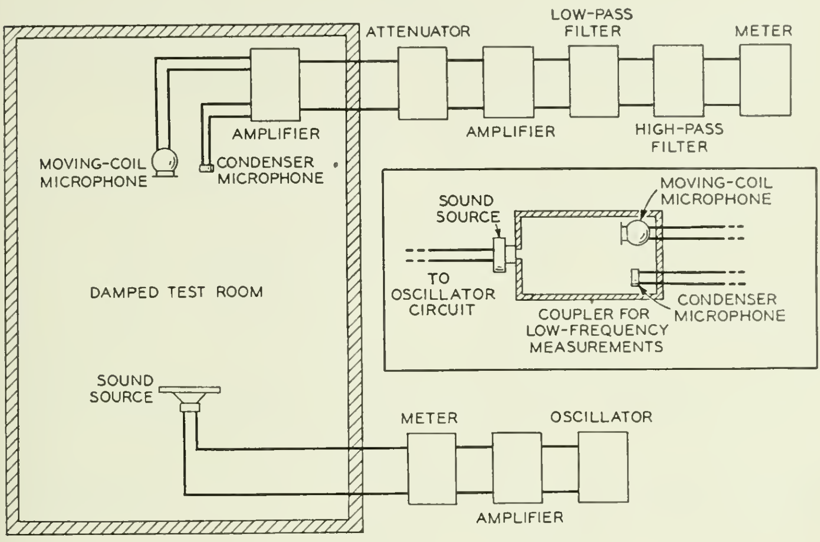


Fig. 9—Response measuring circuit.

the instrument to be tested is substituted at this point and its generated voltage recorded. The response is obtained in terms of decibels below a reference level of one volt per bar of undisturbed sound pressure.

In conclusion we wish to acknowledge our indebtedness to Mr. L. W. Giles and Mr. R. C. Miner of the Bell Telephone Laboratories who have aided us greatly in this work.

APPENDIX A

Theory

Since in most papers on moving coil microphones an equivalent circuit of this instrument is given without deriving it, it does not seem superfluous to indicate here the method of obtaining such a circuit and the calculation of its constants.

In deriving the theory of this instrument it is convenient to speak of the contour pressure response of a microphone.* We shall define the contour pressure response at a certain frequency as the generated open-circuit voltage per bar of uniform pressure over the face of the microphone. On the other hand the field response at a certain frequency is defined as the generated open-circuit voltage per bar of pressure of the undisturbed sound field. The difference between the contour pressure response and the normal incidence field response is caused by diffraction around the microphone, as explained earlier in this paper. We may approximate the normal incidence field response by adding to the contour pressure response the diffraction of the corresponding sphere and circular plate. In the considerations that follow the influence of the screen on the response will be omitted.

To obtain this contour pressure response we assume, then, a uniform pressure over the microphone face. We further require that no wave propagation shall occur within the microphone. Let a small alternating pressure be applied and consider the motion of the system when the pressure is positive. Then the air in the grid holes moves as a whole and imparts an excess pressure to the grid chamber and to the diaphragm. The central portion of the latter moves as a rigid piston. The air volume underneath the diaphragm is compressed, and some air is forced through the coil slot into a very small chamber just in front of the damping ring. We shall assume again that the air in the coil slot moves as a whole. From here the air flows through the acoustic resistance into the larger case chamber. The outside pressure instead of acting upon the diaphragm in the described manner may enter the case through the equalizing tube. On its travel through the tube it is attenuated and its phase is changed. At low frequencies this property of the tube is used to increase the response of the instrument. In the theory that follows, the tube circuit is omitted at first, but its position in the general arrangement will be shown later.

Let us use the following notation for the elements of the vibrating structure:

r_{-1} = equivalent mechanical resistance of all holes in grid,

m_{-1} = equivalent mass of all holes in grid,

n = number of holes in grid,

A_{-1} = area of all holes in grid,

* The term contour pressure response is useful when the microphone has acoustic circuits in front of its diaphragm. Pressure response is a term which has been reserved specifically for the condition where uniform pressure is applied directly to the diaphragm. (See the report of a subcommittee on fundamental sound measurements on the calibration of microphones in the journal of the *Acoustical Society of America*, Vol. 7, April, 1936, p. 301.)

A_0 = effective area of diaphragm,
m_0 = effective mass of diaphragm,
r_0 = mechanical resistance of diaphragm,
S_0 = stiffness of diaphragm,
V_1 = volume of air under diaphragm,
m_2 = equivalent mass in coil slot,
r_2 = equivalent mechanical resistance in coil slot,
A_2 = total area of coil slots,
V_s = volume of air chamber in front of resistance ring,
m_4 = equivalent mass of air in acoustic resistance silk,
r_4 = mechanical resistance to air flow in silk,
A_4 = total area of holes in silk,
V_5 = volume of case,
r_T = equivalent mechanical resistance of tube,
m_T = equivalent mass of air in tube,
A_T = area of tube,
$\dot{x}_{-1,\,0,\,2,\,4}$ = linear velocities,
$x_{-1,\,0,\,2,\,4}$ = linear displacements,
P_0 = atmospheric pressure,
λ = ratio of specific heats for a gas,
$\dfrac{\text{Force}}{\text{Velocity}}$ = mechanical impedance (Force and Velocity are both complex quantities),
$\dfrac{\text{Force}}{\text{Displacement}}$ = stiffness coefficient.

The Lagrangian equations for a system with four independent coordinates can be written in the form:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{x}_n}\right) + \frac{\partial F}{\partial \dot{x}_n} + \frac{\partial V}{\partial x_n} = e_n(t) \qquad (n = -1, 0, 2, 4), \tag{1}$$

in which T is the kinetic energy, F is Rayleigh's dissipation function, V is the potential energy, and $e_n(t)$ is a periodic force. The kinetic energy of the system is

$$T = \tfrac{1}{2}m_{-1}\dot{x}_{-1}^2 + \tfrac{1}{2}m_0\dot{x}_0^2 + \tfrac{1}{2}m_2\dot{x}_2^2 + \tfrac{1}{2}m_4\dot{x}_4^2. \tag{2}$$

Rayleigh's dissipation function becomes

$$\mathrm{F} = \tfrac{1}{2}r_{-1}\dot{x}_{-1}^2 + \tfrac{1}{2}r_0\dot{x}_0^2 + \tfrac{1}{2}r_2\dot{x}_2^2 + \tfrac{1}{2}r_4\dot{x}_4^2, \tag{3}$$

and the potential energy takes the form

$$V = \frac{1}{2}\frac{\lambda P_0}{V_{-1}}(A_{-1}x_{-1} - A_0x_0)^2 + \frac{1}{2}S_0x_0^2 + \frac{1}{2}\frac{\lambda P_0}{V_1}(A_0x_0 - A_2x_2)^2$$
$$+ \frac{1}{2}\frac{\lambda P_0}{V_3}(A_2x_2 - A_4x_4)^2 + \frac{1}{2}\frac{\lambda P_0}{V_5}(A_4x_4)^2. \quad (4)$$

Let

$$a_{-1} = \frac{\lambda P_0}{V_{-1}},$$
$$a_1 = \frac{\lambda P_0}{V_1}, \text{ etc.}$$

Differentiating the above expressions, substituting into (1), writing $\ddot{x} = j\omega\dot{x}$, $x = \frac{\dot{x}}{j\omega}$, and noting that $e_2 = e_3 = e_4 = 0$, we have

$$\left(j\omega m_{-1} + r_{-1} + \frac{a_{-1}A_{-1}^2}{j\omega}\right)\dot{x}_{-1} + \left(\frac{-a_{-1}A_{-1}A_0}{j\omega}\right)\dot{x}_0 = e_1,$$
$$-\frac{a_{-1}A_0A_{-1}}{j\omega}\dot{x}_{-1} + \left(j\omega m_0 + r_0 + \frac{a_{-1}A_0^2 + S_0 + a_1A_0^2}{j\omega}\right)\dot{x}_0 - \frac{a_1A_0A_2}{j\omega}\dot{x}_2 = 0,$$
$$-\frac{a_1A_2A_0}{j\omega}\dot{x}_0 + \left(j\omega m_2 + r_2 + \frac{a_1A_2^2 + a_3A_2^2}{j\omega}\right)\dot{x}_2 - \frac{a_3A_2A_4}{j\omega}\dot{x}_4 = 0,$$
$$-\frac{a_3A_4A_2}{j\omega}\dot{x}_2 + \left(j\omega m_4 + r_4 + \frac{a_3A_4^2 + a_5A_4^2}{j\omega}\right)\dot{x}_4 = 0.$$

If we were to draw the equivalent circuit from these equations we would find that negative stiffnesses are introduced by the different areas through which the air has to flow. In the shunt arms, however, only positive stiffnesses appear. In order to eliminate the negative stiffnesses it is customary to group the shunt stiffness with a negative stiffness and another positive stiffness into a T structure. It is simple to show that this T structure is equivalent to an ideal auto-transformer shunted by a positive stiffness. The turn ratio of the transformer is given by the ratio of two areas. Of course, we may write for the impedance looking into the high side of the auto-transformer

$$Z_H = \left(\frac{A_n}{A_m}\right)^2 Z_L,$$

where

Z_L = impedance in the low side of auto-transformer,
$A_{n,m}$ = areas where $A_n > A_m$.

If a voltage is in series with Z_L it must be multiplied by the ratio of A_n/A_m. If these transformations are carried out we obtain the following equations.

$$\left[j\omega m_{-1}\left(\frac{A_0}{A_{-1}}\right)^2 + r_{-1}\left(\frac{A_0}{A_{-1}}\right)^2 + \frac{a_{-1}A_0^2}{j\omega}\right]\dot{x}_{-1} - \frac{a_{-1}A_0^2}{j\omega}\dot{x}_0 = e_1\frac{A_0}{A_{-1}},$$

$$-\frac{a_{-1}A_0^2}{j\omega}\dot{x}_{-1} + \left[\frac{(a_{-1}+a_1)A_0^2 + S_0}{j\omega} + j\omega m_0 + r_0\right]\dot{x}_0 - \frac{a_1A_0^2}{j\omega}\dot{x}_2 = 0,$$

$$-\frac{a_1A_0^2}{j\omega}\dot{x}_0 + \left[\frac{(a_1+a_3)A_0^2}{j\omega} + j\omega m_2\left(\frac{A_0}{A_2}\right)^2 + r_2\left(\frac{A_0}{A_2}\right)^2\right]\dot{x}_2 - \frac{a_3A_0^2}{j\omega}\dot{x}_4 = 0,$$

$$-\frac{a_3A_0^2}{j\omega}\dot{x}_2 + \left[\frac{(a_3+a_5)A_0^2}{j\omega} + j\omega m_4\left(\frac{A_0}{A_4}\right)^2 + r_4\left(\frac{A_0}{A_4}\right)^2\right]\quad \dot{x}_4 = 0.$$

These equations can be translated into an equivalent circuit in which the effect of the equalizing tube may be inserted as a shunt enabling the impressed force to enter the case and to reach the diaphragm after passing through numerous circuit elements.

The voltage generated in the coil due to its motion in an air gap of a permanent magnet is proportional to $\dot{x}_0$ the velocity of the coil. Hence the expression should be solved for the expression $\frac{e_1}{\dot{x}_0}\frac{(A_0)}{(A_{-1})} = Z$. If l is the length of the wire in the coil, B the flux density in the gap, then the voltage generated in the coil is

$$V = Bl\dot{x}_0 = \frac{Ble_1}{Z}\frac{(A_0)}{(A_{-1})},$$

and since $e_1 = pA_{-1}$ where p is the pressure we have finally

$$\frac{V}{p} = \frac{BlA_0}{Z}\cdot 10^{-8} \text{ volts per bar.}$$

The response in db is equal to

$$\eta = 20\log_{10}\frac{BlA_0}{Z}\cdot 10^{-8}. \tag{5}$$

This quantity when plotted against frequency constitutes the contour pressure calibration. The normal incidence field calibration is found by always adding at any frequency the larger ordinate of the curves representing the diffraction of a sphere and of a flat plate. For convenience these effects shown already in Fig. 6 are given again in Fig. 10 which also compares the theoretical with the experimental response. We observe that the theoretical field calibration is in good agreement with the experimental response.

The meaning of most constants used in evaluating (5) is evident. Some are easily calculated, while others have to be found by measurement. The resistance of the silk is found by allowing air of a certain

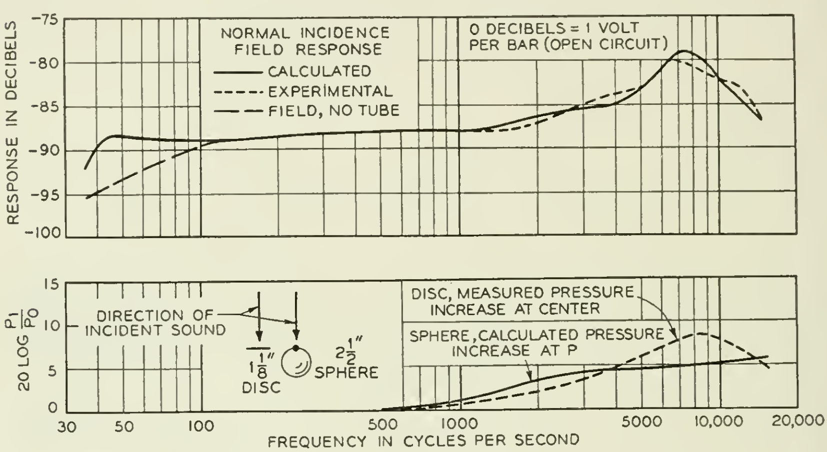


Fig. 10—Field response of 630-A microphone without screen for sound incident normally to diaphragm.

volume velocity to flow through it and by measuring at the same time the pressure drop across the resistance.[1] The mass of the silk is found by impedance measurements.[6] To find the equivalent mass of the coil slot is somewhat difficult since it consists of the mass of the slot plus a certain mass under the dome and under the outer annulus. If the separations between diaphragm and magnet structure are large the problem becomes much simpler since only the mass in the coil slot needs to be considered. The velocity in the slot varies along its width and for any point is given by

$$V = \frac{1}{\mu}\frac{\partial p}{\partial x}\frac{Z(Z - h)}{2}, \tag{6}$$

where Z is the distance from the side wall of the slot to the point in

question, h is the width of the slot, $\frac{\partial p}{\partial x}$ is the pressure gradient and μ is the coefficient of viscosity of air.[7] If m is the mass per unit volume, the kinetic energy for a unit length of the slot and for a unit length along the circumference is

$$\text{K.E.} = \frac{1}{2}\int_0^h \frac{m}{\mu^2}\left[\frac{\partial p}{\partial x}\right]^2 \frac{Z^2(Z-h)^2}{4}\,dZ. \tag{7}$$

The same kinetic energy expressed in terms of the average linear velocity and the effective mass of the whole width is

$$\text{K.E.} = \frac{1}{2}m_e\frac{1}{\mu^2}\left[\frac{h}{12}\right]^2\left[\frac{\partial p}{\partial x}\right]^2. \tag{8}$$

Comparing the integrated expression of (7) with (8) we find that the ratio of the effective mass to the physical mass in the slot is $\frac{6}{5}$, that is

$m_2 = \frac{6}{5}$ mass of two slots.

Knowing the average linear velocity in the slot it is quite simple to calculate the mechanical resistance as

$$r_2 = \frac{24\,\mu\, l\pi D}{h}.$$

If the diameter of the coil is large compared to the air passages then D can be taken to be the diameter of the coil.

The constants r_0 and s_0 can be found from the location and magnitude of the resonant peak when the diaphragm is not damped by any external resistance. In making such measurements it was found that r_0 was a function of frequency. It is sufficient, however, to choose an average value because r_0 is usually small as compared to the resistance of the damping ring. m_0 is again calculated from a consideration of the kinetic energy. If the diaphragm behaves like a simple piston the dome-shaped center portion will have the same velocity at all points. For the annulus we may assume parabolic deflection. The inner region plus the effective mass of the outer annulus make up m_0.

When we consider the grid we again make the assumption that the air in the holes moves like a slug, and that the frictional losses due to the walls can be neglected. Even the impedance due to the effective mass of the slug itself is less important than its radiation impedance. Since the latter is a function of frequency it is necessary to change r_{-1} and m_{-1} for each frequency which is being considered. An account of a

similar problem can be found in I. B. Crandall's "Theory of Vibrating System and Sound" and therefore will not be considered here.[8]

In order to evaluate the constants of the narrow tube used to increase the low-end response we must investigate the discriminant kr. If $|kr|$ lies between the limits $+1$ and $+10$ then the mechanical impedance for a tube of length l and area A_T is given by

$$Z = -\frac{\mu k^2 A_T l}{\left[1 - \frac{2}{kr}\frac{J_1(kr)}{J_0(kr)}\right]},^*$$

where $k = \sqrt{\frac{-\omega\rho i}{\mu}}$, $n = \sqrt{\frac{\omega\rho}{\mu}}$, r the radius of tube and ρ is the density of air. $J_1(kr)$ and $J_0(kr)$ are Bessel's functions of first and zero order respectively with complex argument. Substituting for the values of k and expressing J_0 and J_1 in terms of ber and bei functions we have [9]

$$Z = \frac{i\rho\omega A_T l}{\left[1 - \frac{2}{nr} \times \frac{\text{ber}'\, nr + i\, \text{bei}'\, nr}{-\text{bei}\, nr + i\, \text{ber}\, nr}\right]}.$$

If values of this impedance are plotted it will be found that the resistance and mass components vary again with frequency. It is therefore necessary to use a new value for r_T and m_T for each frequency when the response of the network is calculated.

APPENDIX B

The "pressure calibration" of the miniature condenser microphone is measured on the thermophone, but the field calibration must also be determined very carefully. The field correction to be applied to this thermophone calibration is made up of two factors, (1) the diffraction effect of the microphone and (2) the resonance of the small cavity in front of the diaphragm. The latter has been calculated carefully and checked experimentally (Fig. 11).

The condenser microphone itself was used to determine its own diffraction effect. This is possible because of the verified theoretical law giving the diffraction effect to be a function of the product of the diameter and frequency. For example, the diffraction effect that occurs at 2,000 cycles for a 6-inch disc will occur for a 1-inch disc at 12,000 cycles. Since the diffraction effect of the small condenser microphone is essentially that of a cylinder of the same diameter, it was only necessary to measure the effect of a large cylinder in a frequency

* Reference 8, p. 237.

range where the disturbance caused by the test microphone is negligible. This was accomplished by setting the microphone flush into the face of the obstacle when obtaining the disturbed sound pressure. Then by the simple law given, this diffraction effect was applied to the microphone itself and is shown in Fig. 11. The resultant normal incidence field calibration for the small condenser microphone is shown in the same figure.

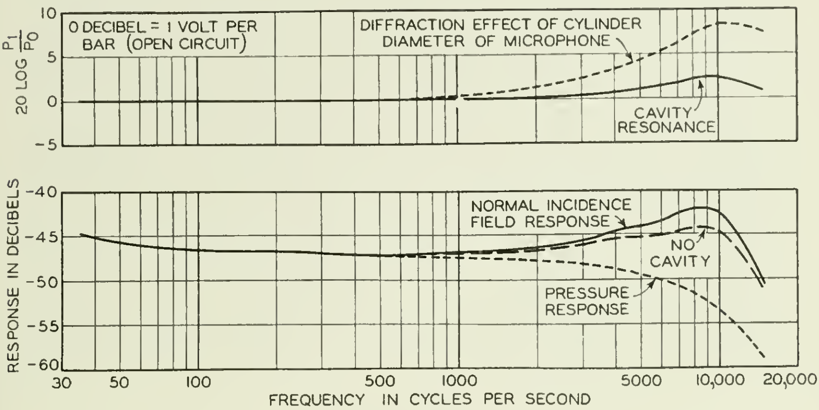


Fig. 11—Response and diffraction of miniature condenser microphone.

References

1. W. C. Jones and L. W. Giles: "A Moving Coil Microphone for High Quality Sound Reproduction," *Jour. Soc. Motion Picture Engineers*, Vol. 17, December 1931, pp. 977–993.
2. Rayleigh: "The Theory of Sound," Vol. 2, 2nd ed., Macmillan & Co., Ltd., 1896, pp. 272–284.
3. Stuart Ballantine: "Note on the Effect of Reflection by the Microphone in Sound Measurements," *Proc. I. R. E.*, Vol. 16, December 1928, pp. 1639–1644.
4. H. C. Harrison and P. B. Flanders: "An Efficient Miniature Condenser Microphone System," *Bell Sys. Tech. Jour.*, Vol. 11, July 1932, pp. 451–461.
5. L. J. Sivian and H. T. O'Neil: "Sound Diffraction by Rigid Circular Plate, Square Plate and Semi-Infinite Screen, *Jour. Acous. Soc. Amer.*, Vol. 3, April 1932, pp. 483–510.
6. P. B. Flanders: "A Method of Measuring Acoustic Impedance," *Bell Sys. Tech. Jour.*, Vol. 11, July 1932, pp. 402–410.
7. Drysdale: "Mechanical Properties of Fluids," D. Van Nostrand Co., Inc., 1924, pp. 113–116.
8. I. B. Crandall: "Theory of Vibrating Systems and Sound," 2nd ed., D. Van Nostrand Co., Inc., 1927, p. 55, pp. 143–149.
9. Dwight: "Bessel Functions for A. C. Problems," *Trans. A. I. E. E.*, Vol. 48, No. 3, July 1929, pp. 812–820.

Oscillations in Systems with Non-Linear Reactance

By R. V. L. HARTLEY

A theoretical study is presented of the properties of a condenser, one plate of which is free to vibrate, when it is included in a circuit containing a generator, the frequency of which is higher than the resonant frequency of the plate and unrelated thereto. It is shown that the plate may be maintained in oscillation at a frequency at or near its mechanical resonance, at the expense of the energy supplied by the generator, provided certain conditions are satisfied. The most favorable condition is one in which the plate is resonant at the frequency of its vibration and the electric circuit is resonant at that of the generator, and at the difference between the generator and plate frequencies, and is anti-resonant at their sum. Under these conditions the generator voltage must exceed a threshold value determined by the impedances and frequencies. This threshold voltage increases as the conditions become less favorable. Expressions are given for the values of the oscillations as functions of the voltage when the threshold is exceeded. When the sum frequency is absent, the energies dissipated at the plate and difference frequencies are in the ratio of the two frequencies.

The oscillations described represent a special case of a class of similar oscillations, all of which depend on the presence of a non-linear reactance. Another special case is a molecular model capable of reproducing the main features of the Raman effect.

Introduction

A TYPE of free oscillation has been found to occur in non-linear coupled systems, which differs from the ordinary type in that the supporting energy is drawn from an alternating sustained source, rather than from a constant source, as in the ordinary vacuum tube oscillator. The particular example of such oscillations to be described here occurs in an electric circuit containing a condenser, one plate of which is elastically supported so as to constitute a mechanically resonant system.

The possibility of such oscillations in a circuit of this kind was discovered [1] in the course of a theoretical study of the possible use of a moving plate condenser as a modulator in a carrier system. Such use was suggested by the fact that, in a condenser, the mechanical force on the plate is proportional to the square of the charge. In this study, it was assumed that a generator of alternating electromotive force of a relatively high carrier frequency was connected with the condenser terminals, and an alternating mechanical force of a relatively low frequency (corresponding to a Fourier component of a speech wave) was applied to the movable plate of the condenser. The plate was not assumed to be resonant. The non-linear relations between

[1] Hartley; *Phys. Rev.*, Vol. 33, p. 289, February, 1929.

charge and mechanical displacement then give rise to currents of the combination, or sideband frequencies. Among the properties of the system which were studied was the reaction of the plate on the mechanical "generator." This was expressed as a mechanical impedance, i.e., the complex ratio of the alternating force to the alternating velocity.

The expression for this mechanical impedance was found to include a negative resistance, which under certain conditions became equal to the positive resistance representing the remainder of the system. It was evident, therefore, that, under these conditions, oscillations of the frequencies involved could persist in the absence of any external driving force on the plate. The existence of such oscillations was first verified experimentally by Mr. E. Peterson. This and a quantitative experimental study of the phenomenon are described in an accompanying paper.[2] Oscillations of the same general type, associated with iron core coils, had been predicted much earlier by the writer and discovered independently by Mr. E. T. Burton.[3]

However, what happened once the threshold condition was passed, was not apparent from this analysis. The answer to this question was found by assuming the existence of the oscillations, computing their values, and determining under what conditions the values are real. Both methods will be employed in what follows.

Representation of the System

In the analysis it will be assumed that, except for the non-linearity associated with the electromechanical coupling, the law of superposition holds throughout. This means that all parts of the system other than the coupling may be represented by linear impedances, of the form

$$Z = R + iX = Ze^{i\varphi}. \tag{1}$$

"Linear," as here used, means that the impedance is independent of the magnitudes of the oscillations.

If then the plate has an alternating velocity of magnitude V_m and phase θ_m, we represent it by $V_m e^{i\theta_m}$. The resultant of all the linear restoring forces may be represented by a force $Z_m V_m e^{i(\varphi_m+\theta_m)}$. All of the quantities involved will, in general, be functions of the frequency. Similarly a current $I_e e^{i\theta_e}$ will be accompanied by a counter electromotive force $Z_e I_e e^{i(\varphi_e+\theta_e)}$, where Z_e is the impedance of the connected electric circuit in series with that of the condenser with its movable plate at rest in the position of zero displacement.

[2] Hussey, L. W. and Wrathall, L. R.; "Oscillations in an Electromechanical System" in this issue of the *Bell Sys. Tech. Jour.*

[3] Peterson, E.; *Bell Laboratories Record*, Feb., 1929, p. 231.

To evaluate the non-linear forces, consider a parallel plate, air condenser of area, A, and normal separation, x_0, one plate of which is fixed and the other of which is free to move under the action of linear restoring forces. Let the movable plate be displaced a distance, x, in a direction to increase the separation, and let a charge, q, be put on the plates. Then it can easily be shown that the static forces tending to oppose the displacements are

$$= Sx + Kq^2, \tag{2}$$

$$e = \frac{q}{C} + 2Kxq, \tag{3}$$

where S is the stiffness of the constraints on the plate; C is the capacitance, in electrostatic units, when x is zero; and

$$K = \frac{2\pi}{A} \tag{4}$$

is a quantity which will be referred to as the constant of non-linearity.

The first terms of (2) and (3) represent components of the forces which were represented above by the mechanical and electrical impedances respectively. Hence only the last terms need be used in expressing the electromechanical coupling.

We shall assume that there is connected in series with the condenser and its associated electric impedance, a generator of negligible internal impedance, which provides an alternating electromotive force, e_g, of amplitude, E_g, and frequency, ω_g, in radians per second. The phase of this generator will arbitrarily be taken as zero.

For the first part of the analysis, we shall assume that there is an alternating force, f_m, exerted on the plate by a "mechanical generator," which has an amplitude, F_m, frequency, ω_m, and phase, ψ_m. We shall investigate the impedance offered to this force in the resulting condition of forced oscillation. In the second part, the mechanical generator will be omitted, and the free oscillations investigated. It is first necessary, however, to determine what frequencies need be considered.

Possible Frequencies

With the system just described there will be developed oscillations, the frequencies of which constitute an infinite series. It will therefore be necessary to introduce limiting assumptions. First let us consider what frequencies may be present in the system. In doing this it must be recognized that the conventional use of complex quantities is not justified when the system is non-linear. This difficulty is avoided and the advantages of the complex exponential notation are retained

if we use the complete exponential expressions for the trigometric functions, since these are real.

Accordingly we shall call the electromotive force of the generator

$$e_g = \frac{E_g}{2}\left[e^{i\omega_g t} + e^{-i\omega_g t}\right]. \tag{5}$$

We shall assume that this is accompanied by an alternating current,

$$i_g = \frac{I_g}{2}\left[e^{i(\omega_g t+\theta_g)} + e^{-i(\omega_g t+\theta_g)}\right]. \tag{6}$$

We shall call the force exerted by the mechanical generator

$$f_m = \frac{F_m}{2}\left[e^{i(\omega_m t+\psi_m)} + e^{-i(\omega_m t+\psi_m)}\right], \tag{7}$$

and the accompanying alternating velocity

$$v_m = \frac{V_m}{2}\left[e^{i(\omega_m t+\theta_m)} + e^{-i(\omega_m t+\theta_m)}\right]. \tag{8}$$

When the corresponding displacements, obtained by integration of (6) and (8), are substituted in the last term of (3), the resulting electromotive force is found to consist of components of frequencies,

$$\omega_s = \omega_g + \omega_m, \tag{9}$$

$$\omega_d = \omega_g - \omega_m, \tag{10}$$

which tend to set up currents at the frequencies of the sidebands.

If such currents flow and we substitute the charges associated with them, together with that from (6), in the last term of (2), we find, in the force on the plate, components of frequency ω_m, and a variety of other frequencies including zero, i.e., a steady force. If these produce displacements which are again substituted in (2), and the process is continued, we arrive finally at the entire series of frequencies given by $m\omega_g \pm n\omega_m$, where m and n are integers.

We shall now introduce the limiting assumption that the plate is resonant at or near ω_m, and not at any other frequency. The impedance at that frequency will then be small and the response to the driving force at that frequency relatively large. At the frequencies of all the other components of the force the mechanical impedance will be relatively very high; and we will not be making a violent assumption if we say that it is so high that the velocities of response at all the other frequencies are negligible. [There may be some response to the steady force, consisting of a slight change in the position of equilibrium about which the vibrations occur. This can be taken care of by saying that the coefficients in (2) and (3), while constant for any

particular condition of sustained oscillation, vary slightly with the magnitude of the oscillations.] With this assumption the frequencies of the components of the electromotive force reduce to $\omega_g \pm n\omega_m$.

If the electric circuit is resonant at any of these frequencies, we may as above neglect the currents at other frequencies. In the absence of any resonance, if the constant of non-linearity, K, is sufficiently small, the amplitudes will decrease so rapidly with increasing n that we may neglect all for which n is greater than unity. In the interests of simplicity we shall make all of these assumptions. We have then in addition to i_g and V_m the currents,

$$i_s = \frac{I_s}{2}\left[e^{i(\omega_s t+\theta_s)} + e^{-i(\omega_s t+\theta_s)}\right], \tag{11}$$

$$i_d = \frac{I_d}{2}\left[e^{i(\omega_d t+\theta_d)} + e^{-i(\omega_d t+\theta_d)}\right]. \tag{12}$$

Forced Oscillations; Impedance Solution

We wish to set up the equations of motion in terms of the applied forces, the velocities and currents at the various frequencies, and the properties of the system, as expressed in terms of its linear impedances and the constant of non-linearity, K. For each frequency, we equate whatever applied force there may be to the sum of the restoring forces due to the system. These consist of a component given by the product of the velocity or current by the impedance for the particular frequency, and other terms due to the combination of pairs of the other frequencies. To find these latter components, we integrate (6), (8), (11) and (12) with respect to time, substitute the resulting displacements in the last terms of (2) and (3), and select the components of the four significant frequencies, for insertion in their appropriate equations. Once these components are obtained, we may, since the remainder of the system is linear, safely revert to the conventional use of exponentials, so that the factor $e^{i\omega t}$ may be divided out for each equation. The final result is

$$Z_g I_g e^{i(\theta_g+\varphi_g)} + \frac{KV_m I_s}{\omega_m\omega_s} e^{i(\theta_s-\theta_m)} - \frac{KV_m I_d}{\omega_m\omega_d} e^{i(\theta_m+\theta_d)} = E_g, \tag{13}$$

$$Z_m V_m e^{i(\theta_m+\varphi_m)} + \frac{KI_g I_s}{\omega_g\omega_s} e^{i(\theta_s-\theta_g)} + \frac{KI_g I_d}{\omega_g\omega_d} e^{i(\theta_g-\theta_d)} = F_m e^{i\psi_m}, \tag{14}$$

$$Z_s I_s e^{i(\theta_s+\varphi_s)} - \frac{KI_g V_m}{\omega_g\omega_m} e^{i(\theta_g+\theta_m)} = 0, \tag{15}$$

$$Z_d I_d e^{i(\theta_d+\varphi_d)} + \frac{KI_g V_m}{\omega_g\omega_m} e^{i(\theta_g-\theta_m)} = 0. \tag{16}$$

These equations are of the second degree and so are not so simple of solution as are the linear equations of circuit theory which they formally resemble. We note, however, that if, in the last three, we assume I_g and θ_g to be constant, they become linear. We may therefore solve them as linear equations, with this assumption, provided we bear in mind that the resulting impedances will not be linear unless the oscillations are so small that the second and third terms of (13) can be neglected compared with the first.

Let us make this assumption and explore the properties of the resulting linear system represented by (14), (15) and (16). If we calculate $V_m e^{i\theta_m}$ and take the ratio $F_m e^{i\psi_m}/V_m e^{i\theta_m}$, this will be the analog of the impedance of an analogous electric circuit as measured in the mesh corresponding to vibration of the plate at frequency ω_m. This ratio, which we shall call Z_m', may be thought of as the mechanical impedance of the plate when the circuit is activated by the electrical generator. Following circuit theory, as applied to vacuum tubes, let us call Z_m' the active impedance of the plate, and Z_m the passive impedance. The value of the active impedance, when expressed in terms of resistances and reactances, is found to be

$$Z_m' = (R_m + iX_m) + \frac{K^2 I_g^2}{\omega_g^2 \omega_m \omega_s} \cdot \frac{R_s - iX_s}{Z_s^2} + \frac{K^2 I_g^2}{\omega_g^2 \omega_m \omega_s} \cdot \frac{-R_d - iX_d}{Z_d^2}. \quad (17)$$

We see that the active impedance differs from the passive impedance by two terms, each of which represents the effect of the impedance of the electric circuit at one of the side frequencies. The second term of (17), which depends on the impedance at the sum frequency, is identical in form with the impedance added to an electric circuit,[4] at a frequency, ω, by a transformer of mutual inductance, M, provided that

$$M^2\omega^2 = \frac{K^2 I_g^2}{\omega_g^2 \omega_m \omega_s}; \quad (18)$$

the impedance of the secondary circuit is equal to Z_s; and the reactances of the primary and secondary windings are included in X_m and X_s, respectively. The third term which depends on the impedance of the electric circuit at the difference frequency, is similar except that the effective resistance is negative.

It is this negative resistance which makes possible the type of free oscillations here described. To interpret it, let us start with the small

[4] Bush, V.; "Operational Circuit Analysis," John Wiley & Sons, Inc., 1929, p. 50, Eq. (66).

applied force, f_m, acting on the plate, with the voltage of the electric generator zero. The velocity of the plate vibration is then determined by its passive impedance alone. Let us assume for the time being that the impedance, Z_s, of the electric circuit at the sum frequency is infinite, so that its effect on the active impedance of the plate disappears. Let us make φ_d equal to φ_m, and gradually increase the generator voltage. As I_g^2 increases, the negative impedance increases, the total impedance decreases and the velocity, V_m, increases. This condition is analogous with the behavior of the input impedance of a regeneratively connected amplifier when the plate current is progressively increased from zero. At a threshold value of I_g, the net impedance becomes zero and the velocity infinite. This means that a finite velocity can exist for an infinitesimal driving force, that is, the oscillations, once started, are self-sustaining, even in the absence of any sustained driving force, f_m, at the mechanical frequency.

If we make the electric impedance, Z_d, at the difference frequency infinite, all the resistances are positive; so sustained oscillations cannot occur, in a dissipative system, in the absence of current at the difference frequency. If both side frequencies are present, so that Z_s and Z_d are both finite, sustained oscillations are still possible provided the impedance at the sum frequency is not too small compared with that at the difference frequency. The presence of current at the sum frequency always increases the critical value of the current at the generator frequency.

We may also compute the active impedance of the electric circuit at the side frequencies, on the same assumption as to the constancy of the current of generator frequency as was made in deriving (17). To do this, we remove the mechanical generator, making the right member of (14) zero, and insert low measuring voltages of frequencies ω_s and ω_d in the right members of (15) and (16), in turn. In each case we compute the ratio of this voltage to the accompanying current. If we think of each frequency as being the analog of a mesh in an electric circuit, we note that the mesh corresponding to the mechanical frequency is coupled to both of the side frequencies; but the latter are not directly coupled to each other. If the mutual impedances, which depend on I_g, are small enough, we may, for a generator at the sum frequency, neglect the third term of (14), which represents the effect of the loosely coupled difference frequency mesh, compared with the first. The active impedance at the sum frequency then becomes

$$Z_s' = (R_s + iX_s) + \frac{K^2 I_g^2}{\omega_g^2 \omega_m \omega_s} \cdot \frac{R_m - iX_m}{Z_m^2} \cdot \qquad (19)$$

If the third term of (14) is not neglected we must replace Z_m in (19), by the first and third terms of (17), that is, by the impedance of the mechanical frequency mesh, as modified by its coupling with the difference frequency mesh.

Similarly, when the measuring generator is of the difference frequency, we get

$$Z_d' = (R_d + iX_d) + \frac{K^2 I_g^2}{\omega_g^2 \omega_m \omega_s} \cdot \frac{-R_m - iX_m}{Z_m^2}, \qquad (20)$$

where Z_m is to be replaced by the first and second terms of (17), if the second term of (14) is not neglected.

The active impedance at the difference frequency (20) contains a negative resistance similar to that which appeared at the mechanical frequency (17). In fact, if the passive impedance, Z_s, at the sum frequency is infinite, the expressions for the two active impedances are symmetrical. The active impedance at the sum frequency contains only positive resistances, except in so far as the resistance of the mechanical mesh is made negative by its coupling with the difference mesh. This serves to emphasize the fact that the presence of current of the difference frequency is essential to the oscillations, while that of current of the sum frequency tends to make their production more difficult.

Free Oscillations

In the above considerations it was assumed that the amplitudes at all of the new frequencies were small compared with that at the generator frequency. While this assumption permits us to compute the threshold conditions for the starting of free oscillations, it is violated as soon as the oscillations become appreciable. In order to find out what happens once the threshold is passed it is necessary to solve the second degree equations (13) to (16) when F_m is made zero. The presentation of this solution will be simplified by considering first the case where the sum frequency is eliminated and then the effect of its presence on the simpler solution.

The elimination of the sum frequency is accomplished by making Z_s infinite and I_s zero. This makes the second terms of (13) and (14) zero, and makes (15) indeterminate. We are left then with (13), (14), as modified, and (16). The equations for the mechanical and difference frequencies are now symmetrical. In order to solve these equations we express the exponentials in terms of sines and cosines and equate the real and imaginary parts separately. In the equations derived from (14) and (16) we transpose the second term in each equation to the right member. For each pair we divide the equation containing sines by

that containing cosines and obtain a relation between the angles involved. We square each equation of a pair, and add them to obtain a relation between the magnitudes of velocities and currents, the impedances, and the frequencies. From these it follows that I_g is a constant. By means of these relations the equations derived from (13) may be reduced to a form where the only variables are E_g, V_m and θ_g, and the constant, I_g, appears only as a divisor of E_g. These equations are then squared and added to give an equation which determines V_m. The final solution takes the form

$$\varphi_m = \varphi_d, \tag{21}$$

$$I_g = \frac{\omega_g}{K}\left[Z_m\omega_m Z_d\omega_d\right]^{1/2}, \tag{22}$$

$$V_m = \frac{\omega_g}{K}\left[Z_d\omega_d Z_g\omega_g\left(-\cos(\varphi_m+\varphi_g) \pm \left\{\left(\frac{E_g}{Z_g I_g}\right)^2 - \sin^2(\varphi_m+\varphi_d)\right\}^{1/2}\right)\right]^{1/2}, \tag{23}$$

$$I_d = \left[\frac{Z_m\omega_d}{Z_d\omega_m}\right]^{1/2} V_m, \tag{24}$$

$$\theta_g = \varphi_m + \alpha \pm \frac{\pi}{2}, \tag{25}$$

where

$$\cos\alpha = \frac{Z_g I_g}{E_g}\sin(\varphi_m+\varphi_g), \tag{26}$$

and the sign in (25) is so chosen that

$$-\frac{\pi}{2} < \theta_g < \frac{\pi}{2}, \tag{27}$$

and

$$\theta_m + \theta_d = \alpha + \pi \pm \frac{\pi}{2}, \tag{28}$$

where the same sign is to be taken for $\pi/2$ as in (25).

The nature of the variation represented by (23) is shown in Fig. 1, which is taken from the accompanying experimental paper.[2] Here the amplitude, V_m/ω_m, of the plate displacement is plotted against the generator voltage, E_g, for the case of exact resonance and for one involving a slight departure from resonance.

Let us interpret these results physically. The phase angles in (21) depend only on the physical constants of the system and the frequencies of the oscillations. This equation, therefore, determines at what frequencies oscillations may occur provided the other conditions are

satisfied. Thus the ratio of reactance to resistance must be the same for the plate at ω_m and for the electric circuit at ω_d. This condition is satisfied if each is resonant at its particular frequency, but resonance is not a necessary condition. All that is necessary is that there be a pair of frequencies, whose sum is equal to that of the electric generator, for which the impedances have the same phase angle. If there are an electric and a mechanical resonance such that the sum of the resonant frequencies is nearly equal to the generator frequency, and there is a marked difference in the sharpness of the two resonances, then the oscillations will fall closer to the sharper resonance. This is due to the fact that the phase angle of the impedance changes more rapidly with frequency in the neighborhood of a sharp resonance.

From (22) we see that the amplitude of the current at the generator frequency depends only on this frequency, the constants of the system, and the new frequencies. It is independent of the amplitude of the

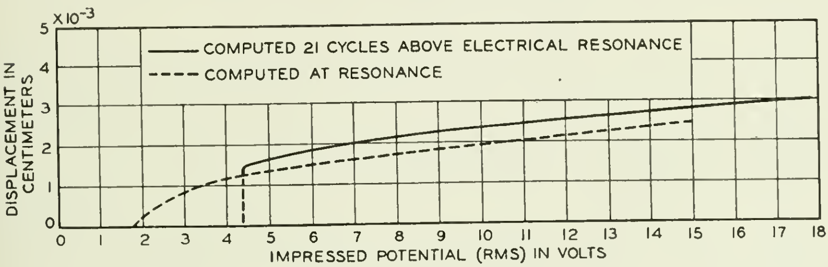


Fig. 1—Alternating displacement of plate as a function of generator voltage.

generator voltage, of the amplitudes of the new frequencies, and of the impedance of the electric circuit at the generator frequency. This equation, while it tells us what happens when the oscillations are present, tells us nothing about the conditions for their existence. These are to be found by noting under what conditions the expression (23) for the amplitude at the new frequency, ω_m, is real. We have two cases to consider which are determined by the sign of $\cos(\varphi_m + \varphi_g)$.

Assume first that this is positive, as would be the case if the plate is resonant at ω_m and there is any dissipation at ω_g, such as would be caused by resistance in the electric circuit. The first term in (23) is negative and V_m can be real only if the second term exceeds it in absolute value. This condition reduces to

$$E_g > Z_g I_g = \frac{Z_g \omega_g}{K} [Z_m \omega_m Z_d \omega_d]^{1/2}. \tag{29}$$

This shows that there is a threshold value of the generator voltage,

above which the new oscillations are possible. (It is found to agree with that obtained by the negative resistance method.) Moreover, this value is that which is just necessary to maintain an electric current, of the generator frequency, in the absence of the new frequencies, with an amplitude equal to the constant amplitude that exists in the presence of the new frequencies. For values of E_g large compared with the threshold value, the amplitudes of the new frequencies increase nearly as the square root of the amplitude of the generator voltage.

In the special case of resonance at both ω_m and ω_d, Z_m and Z_d tend to be small and so from (24) the threshold voltage is correspondingly small. This therefore is a particularly favorable condition for the production of the oscillations.

The case where cos $(\varphi_m + \varphi_g)$ is negative occurs when all of the three impedances are predominantly reactive, the reactances being all of the same sign. The first term of (23) is then positive and V_m will be real if the second term is positive, as it will be if

$$E_g > Z_g I_g |\sin (\varphi_m + \varphi_g)|. \tag{30}$$

For this case, then, the threshold amplitude of the generator voltage may be much less than that required to maintain the current at the constant amplitude, I_g, in the absence of the new frequencies.

In the extreme case where there is no dissipation and the phase angles of the impedances are all $\pm \pi/2$, the threshold voltage reduces to zero and so sustained oscillations are possible in the absence of any generator. (23) and (24) then reduce to forms symmetrical with (22). This means that for such a system the frequencies would be determined by the constants of the system and the amount of energy present, since this would limit the possible amplitudes.

There is some question as to the sign to be given to the inner radical in (23). When cos $(\varphi_m + \varphi_g)$ is positive the plus sign must be used. When it is negative the plus sign must be used if E_g is greater than $Z_g I_g$. If E_g is between this and the threshold given by (30), either sign gives a real value for the amplitude. When the sign is negative the amplitude decreases with increasing voltage, which appears to be an unstable condition.

Regarding the phases, the condition represented by (27) is imposed because the energy flow must be from the generator to the circuit. Only the sum of the phases of the new oscillations is determined. Their individual values depend on the starting conditions, just as does the phase of a pendulum clock.

One more result may be of interest. This is the relative rates at which energy is dissipated at the two new frequencies. If P_m and P_d

are the powers corresponding to the two frequencies we have

$$\frac{P_d}{P_m} = \frac{I_d^2 Z_d \cos \varphi_d}{V_m^2 Z_m \cos \varphi_m} = \frac{\omega_d}{\omega_m}. \tag{31}$$

Thus the rate of energy dissipation is in the ratio of the frequencies.

Effect of Sum Frequency

The more general case where the sum frequency is also present calls for the solution of (13) to (16) as they stand except for F_m being zero. This may be done by substituting the values of I_s and θ_s from (15), and those of I_d and θ_d, from (16), in (13) and (14), and proceeding in a manner similar to that used above. The results take the form

$$\varphi_m = \gamma, \tag{32}$$

where

$$\tan \gamma = \frac{Z_s \omega_s \sin \varphi_d + Z_d \omega_d \sin \varphi_s}{Z_s \omega_s \cos \varphi_d - Z_d \omega_d \cos \varphi_s}, \tag{33}$$

and the signs of sin γ and cos γ are determined by the numerator and denominator of (33) respectively;

$$I_g = \frac{\omega_g}{K}\left[\frac{Z_m \omega_m Z_d \omega_d}{a}\right]^{1/2}, \tag{34}$$

where

$$a = \left[1 + \left\{\frac{Z_d \omega_d}{Z_s \omega_s}\right\}^2 - 2\frac{Z_d \omega_d}{Z_s \omega_s}\cos(\varphi_d + \varphi_s)\right]; \tag{35}$$

$$V_m = \frac{\omega_m}{K}\left[\frac{Z_d \omega_d Z_g \omega_g}{b}\left(-\cos(\delta + \varphi_g) \pm \left[\left(\frac{E_g}{Z_g I_g}\right)^2 - \sin^2(\delta + \varphi_g)\right]^{1/2}\right)\right]^{1/2}, \tag{36}$$

where

$$b = \left[1 + \left(\frac{Z_d \omega_d}{Z_s \omega_s}\right)^2 + 2\left(\frac{Z_d \omega_d}{Z_s \omega_s}\right)\cos(\varphi_d - \varphi_s)\right]^{1/2}, \tag{37}$$

and

$$\tan \delta = \frac{Z_s \omega_s \sin \varphi_d + Z_d \omega_d \sin \varphi_s}{Z_s \omega_s \cos \varphi_d + Z_d \omega_d \cos \varphi_s}; \tag{38}$$

$$I_d = \left[\frac{Z_m \omega_d}{Z_d \omega_m a}\right]^{1/2} V_m; \tag{39}$$

$$I_s = \frac{Z_d}{Z_s} I_d; \tag{40}$$

$$\theta_g = \delta + \alpha' \pm \frac{\pi}{2}, \tag{41}$$

where

$$\cos \alpha' = \frac{Z_g I_g}{E_g} \sin (\delta + \varphi_g); \tag{42}$$

$$\theta_m + \theta_d = \alpha' + \pi \pm \frac{\pi}{2} + (\delta - \varphi_d); \tag{43}$$

$$\theta_s - \theta_m = \alpha' \pm \frac{\pi}{2} + (\delta - \varphi_s), \tag{44}$$

where the sign of $\pi/2$ is again to be chosen so as to satisfy (27).

Corresponding to (21) we have (32). If the mechanical motion involves any dissipation, the mechanical resistance, $Z_m \cos \varphi_m$, must be positive, and since Z_m is positive by definition, $\cos \varphi_m$ must be positive. This means that (32) can be satisfied only if the denominator of (33) is positive. Hence oscillations can occur only if

$$\frac{Z_d \omega_d}{Z_s \omega_s} < \frac{\cos \varphi_d}{\cos \varphi_s} \cdot \tag{45}$$

This relation can hold when Z_d, the impedance at the difference frequency, is infinite, only if φ_s is $\pm\, \pi/2$, that is, if there is no dissipation at the sum frequency.

To investigate the relative rates of dissipation at the sum and difference frequencies, we find the ratio of the powers P_s and P_d, associated with them.

$$\frac{P_s}{P_d} = \frac{Z_d \cos \varphi_s}{Z_s \cos \varphi_d} < \frac{\omega_s}{\omega_d} \cdot \tag{46}$$

Thus the ratio is always less than the ratio of the frequencies and approaches it only as the limiting condition for oscillations is approached.

A discussion of all possible values of impedance and phase angle at the two side frequencies would be too involved to go into here. The special case of resonance at both frequencies is, however, of some interest since a given current is then accompanied by a maximum of dissipation. It also provides that ω_m coincides with the mechanical resonance, where Z_m is much smaller than for nearby frequencies. Since Z_m enters into the expression for the threshold force, this condition is particularly favorable for the occurrence of oscillations. When we make φ_d and φ_s zero we see from (45) that the impedances, now pure resistances, must be such that

$$\frac{Z_d \omega_d}{Z_s \omega_s} < 1. \tag{47}$$

(35) now becomes

$$a = 1 - \frac{Z_d\omega_d}{Z_s\omega_s}. \tag{48}$$

From (34) it is evident that when Z_s is finite the constant current of the generator frequency, and so also the threshold voltage, are greater than when it is infinite. Thus the presence of current of the sum frequency makes the conditions for oscillation more exacting. As we approach the limiting impedance ratio, where the powers approach the ratio of their frequencies, the threshold voltage approaches infinity, and the probability of oscillations approaches zero.

The relative powers at the difference frequency and at the mechanical frequency are now given by

$$\frac{P_d}{P_m} = \frac{\omega_d}{\omega_m\left(1 - \dfrac{Z_d\omega_d}{Z_s\omega_s}\right)}. \tag{49}$$

The presence of a finite impedance at the sum frequency increases this ratio over that of the frequencies. For the limiting condition of oscillations it approaches infinity, the amplitude at the difference frequency then becoming infinite and that at the mechanical frequency remaining finite.

From these results it appears that proportionality between power and frequency is a limiting case which occurs only under the conditions which are most and least favorable for the existence of oscillations. We should, therefore, expect to find it only under the favorable conditions where the transformation of energy is from a higher to a pair of lower frequencies.

Effect of Other Frequencies

In the interests of simplicity the above treatment was limited to the case where all but four frequencies are suppressed by high impedances. Such a limitation is not, however, essential to the production of oscillations. In fact, as many as desired of the series $m\omega_g + n\omega_m$ may be produced by the proper choice of impedances and the use of high enough voltages, provided, of course, the apparatus can withstand the stresses involved. In general, the presence of certain frequencies will be favorable to oscillations and that of others unfavorable.

Summary

By way of summary, then, it is possible to maintain a movable condenser plate in sustained oscillation by applying to the condenser an

alternating electromotive force of an unrelated higher frequency, provided that the impedances of the system at these two frequencies and their various combinations satisfy certain relations, and the applied electromotive force exceeds a threshold value. When the oscillations are negligible at all frequencies except these two and their sum and difference, the most favorable condition, lowest threshold voltage, occurs when the plate vibrates at its resonant frequency, and the electric circuit is resonant at the applied frequency and at the difference frequency, and anti-resonant at the sum frequency. Once the oscillations start, the current of the applied frequency remains constant with increasing voltage. Under the most favorable conditions the rates of energy dissipation at the plate and difference frequencies are in the ratio of the frequencies.

Other Applications; Raman Effect

While in the case considered above the production of oscillations was associated with a particular type of non-linearity, the application of the principle is much more general. Here the non-linearity occurs in what might be called a mutual stiffness, serving to couple two degrees of freedom. It is not essential, however, that the non-linearity occur in a mutual impedance nor that the impedance be of the stiffness or negative reactance type. So long as the connected system is such as to provide the proper impedances, oscillations may occur in connection with any non-linear reactance.

A non-linear reactance, as here used, may be defined as any energy-storing element in which the coefficient of inertia is a function of the velocity, or that of stiffness is a function of the displacement, or any mechanical, electrical or electromechanical analog, of such an element. For a non-linear inertia, as in an iron core inductance coil, however, the power varies inversely as the frequency; instead of directly as for a non-linear stiffness.

A special case, in which one of the new frequencies is an exact submultiple of the driving frequency, has been studied by a number of workers from Rayleigh [5] down to Pedersen.[6]

Another special case may be of some interest to physicists because it provides a model of the Raman effect. The transition from the condenser to the molecular model will be made in two steps. For the first suppose that instead of making the resonant mechanical member one plate of a condenser, we attach the moving part to a point on its support by an elastic string under tension, the direction of the string

[5] Rayleigh; "Theory of Sound," *Sec. Ed.*, Vol. 1, p. 81.
[6] Pedersen; *Jr. Acous. Soc. Amer.*, Vol. VI, 4, p. 227, April, 1935.

being parallel to the direction of vibration. Suppose now that at some point along the string we apply an alternating mechanical force, acting normal to the string, through the medium of a mechanical structure, which as viewed from the string may be represented by a linear mechanical impedance. This structure prevents motion of its point of attachment to the string, in the direction of the string.

If now we analyze the forces and motions into their components in the direction, x, of the motion of the vibrating member and that, y, of the applied force, we find that, to a first approximation, the relations connecting them are identical with those used above for the condenser, provided we identify the force and velocity of the vibration in the x direction with those of the condenser plate and those of the point of attachment of the string, in the y direction, with the electromotive force and current in the electric circuit associated with the condenser. Such a structure can therefore produce oscillations of the sort described, provided the mechanical impedance of the driving structure has the proper values at the sum and difference frequencies.

Suppose now we have a molecule which we assume to be rigid with the exception of one atom, which is bound to it by a pair of electrons. Let the attached atom correspond to the plate, the relatively heavy molecule to the support and the electrons to the point of application of the driving force. Let the forces of electrostatic attraction between the electrons and the atom, and between the electrons and the center of the molecule, correspond to those due to the tense strings. Let the other static forces between the atom and the molecule correspond to the stiffness of the plate. For small displacements these forces may be assumed to vary linearly with distance, and so be capable of representation by constant coefficients of stiffness which correspond to the elasticities in the mechanical system. The applied external force is that exerted on the electrons by that component of the incident light which is normal to the line through the centers of the undisplaced particles. The mechanical impedance of the electrons for motion in the direction of the applied force corresponds to that of the structure through which the force is applied. This impedance includes the effects of any elastic constraints the rest of the molecule may exert on the electrons in this direction; of the electromagnetic mass of the electrons, which may be affected by the reactions of neighboring molecules; and of the dissipation of energy as radiation or by transfer to neighboring molecules.

Unlike other classical models of the Raman effect, this one provides for the persistence of the difference line, and the disappearance of the sum line, at low temperatures. It also provides that the intensity of

the lines should depend on the probability that the force exerted by the incident radiation on the electrons of a randomly chosen molecule exceed a threshold value which is determined by the condition of its neighbors. The apparent smallness of this probability would explain the observed weakness of the Raman lines.

It would seem that this threshold, and the probability of its being exceeded, might prove helpful in interpreting the energy threshold and transition probability which are used in wave mechanics.

Acknowledgment

In conclusion, I wish to express my thanks to those of my colleagues who by discussions and suggestions have contributed to the preparation of this paper, and in particular to Mr. L. A. MacColl for some valuable suggestions on the mathematical side.

Oscillations in an Electromechanical System

By L. W. HUSSEY and L. R. WRATHALL

Experimental results are given on an oscillating electromechanical system in which, under a single frequency impressed electromotive force, mechanical vibrations are sustained at a frequency near the resonant frequency of the mechanical system and electrical oscillations at the difference between the frequency of the mechanical vibration and that of the impressed force.

The system is the one studied analytically by R. V. L. Hartley in an accompanying paper. Its performance conforms to the principal operating features predicted in his analysis.

IN AN accompanying paper [1] an analytic investigation is made of a system involving a non-linearity in the coupling between an electrical and a mechanical system. The electro-mechanical system under discussion is, in its simplest form, a condenser, with one plate sharply resonant mechanically, a generator, and an impedance, all connected in series. If the charge on the condenser is q, there will be a force on the mechanical system proportional to q^2. While the mechanical system and the electrical system involved are individually linear, there is a non-linearity in this electrostatic coupling, and hence the possibility exists of mechanical and electrical vibrations at other frequencies than the impressed frequency. On this basis the possibility of the generation of a mechanical vibration, *not at a harmonic of the impressed electromotive force*, and electrical currents at the difference between the frequency of the mechanical vibration and that of the impressed electromotive force was predicted by the analysis.

That the phenomenon discussed can occur was first verified by Mr. Eugene Peterson. A condenser microphone was given a mechanical resonance at 600 cycles per second, by cementing a small metal ball to the center of the diaphragm. An alternating electromotive force at 2200 cycles per second was impressed and the system given a series resonance at the difference frequency, 1600 cycles per second, by means of an inductance. When the impressed voltage was increased beyond a critical value mechanical vibrations suddenly built up and current of the difference frequency, larger in amplitude than the current of the impressed frequency, appeared in the electrical system. The same result was obtained using a prong of a tuning fork as the vibrating plate.

[1] "Oscillations in Systems with Non-Linear Reactance" by R. V. L. Hartley, in this issue of the *Bell Sys. Tech. Jour.*

These tests, while they exhibited the most important characteristic predicted, did not give any other check on the validity of the mathematical results because the systems differed so greatly from that assumed by Mr. Hartley. To obtain simple results in the theoretical discussion it was found necessary to make some severely restricting assumptions. The mechanical system was assumed rigid except to motion at frequencies near the resonant frequency. Similarly the electrical system was assumed to have infinite impedance to all frequency components except that impressed and the difference frequency between the impressed and the mechanical. Thus the only currents and velocities present were of the frequencies (in radians per second), ω_m (mechanical), $\omega_d = \omega_g - \omega_m$ (difference), ω_g (impressed).

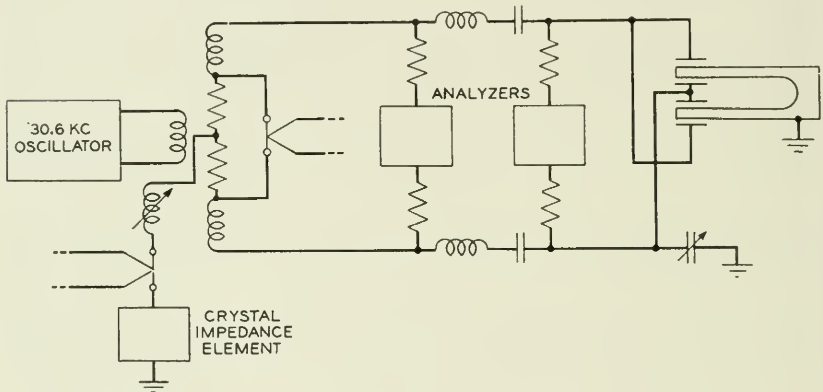


Fig. 1—Circuit diagram.

Under ordinary conditions a non-linear system, such as this one, involving two frequency components (ω_g and ω_m) would have, as modulation products, currents and velocities of all the possible combination frequencies ($r\omega_d + s\omega_m$, r, $s = 0, \pm 1, \pm 2, \cdots$). There would be dissipation of energy at each of these frequencies. These components (other than the three of interest) are the ones which must be suppressed if the system is to be a good approximation to the assumed one.

In order to satisfy the above conditions the circuit of Fig. 1 was constructed.[2] The use of the parallel system instead of a simple series circuit had several advantages. The forces on the tuning fork were so balanced that any constant displacement was avoided. Since the tuning fork had a very sharp resonance [3] the mechanical system was

[2] The tuning fork with condenser plates was designed by Mr. W. A. Marrison.
[3] The damping effect of the air was avoided by operating the fork in a vacuum.

a close approximation to the assumed one. On the electrical side the parallel system had the advantage that the current of the impressed frequency (ω_g) flowed around the outside while the sum ($\omega_g + \omega_m$) and difference ($\omega_d = \omega_g - \omega_m$) frequency components flowed through the mid-branch. The sum frequency component was the most difficult one to suppress. This could be done in the parallel system by means of a piezo-crystal impedance element,[4] tuned to ω_d without at the same time putting in a high impedance to ω_g. This piezo-crystal element had so sharp a resonance that, while its impedance at resonance (30 kilocycles per second) was only 125 ohms, it was about 60,000 ohms only 1000 cycles per second away from resonance. Thus the very high impedance to the sum frequency was obtained. There were, however, some modulation products which flowed around the outside with the impressed current. Only the impedance of a simple tuned circuit was available around this circuit so any product near in frequency to the impressed frequency would not face a very high impedance. The nearest one was $\omega_g - 2\omega_m$. This was not as completely suppressed as the other unwanted products but changing the impedance so that this component was considerably different in magnitude had no appreciable effect on the other electrical components. While this circuit gives a good approximation to the hypothetical one, the result is a highly critical system demanding very fine adjustment and a highly stable generator, since a very slight frequency change has an appreciable effect on the impedances presented by the very sharply tuned circuits.

The measurement of impressed current was made by the thermocouple across small resistors in the input transformer and checked by a current analyzer.[5] Corrections were made in the results when necessitated by the presence of the current component of frequency $\omega_g - 2\omega_m$. The corresponding voltage was obtained by measuring the current of that frequency through the large resistors R_1 by means of a current analyzer. The current of the difference frequency was measured by a thermocouple in the mid-branch. Measurement was also made of the voltage of frequency $\omega_g - 2\omega_m$ across the fork, by measuring the corresponding current through the resistances R_2. There was no simple means available for measuring the mechanical amplitude but it could be obtained from this measurement of $\omega_g - 2\omega_m$ voltage as will be explained later.

On Figs. 2 and 3 are shown curves computed from equations 28, 29 and 30 of the preceding paper for two cases, and the results of the

[4] Designed by Mr. W. P. Mason.

[5] See A. G. Landeen: "Analyzer for Complex Electric Waves," *Bell Sys. Tech. Jour.*, April, 1927.

experiment. The two cases computed are for the system operating with ω_d and ω_m exactly the frequencies of electrical and mechanical resonance and for operation with ω_d 21 cycles higher than resonance. In the latter case ω_m is determined by the requirement—noted by Mr. Hartley—that the phase angles of the impedances be equal. The actual change in ω_m is only a fraction of a cycle. The two cases are given because it was not possible, with the equipment available, to determine ω_d with sufficient accuracy to distinguish between them.

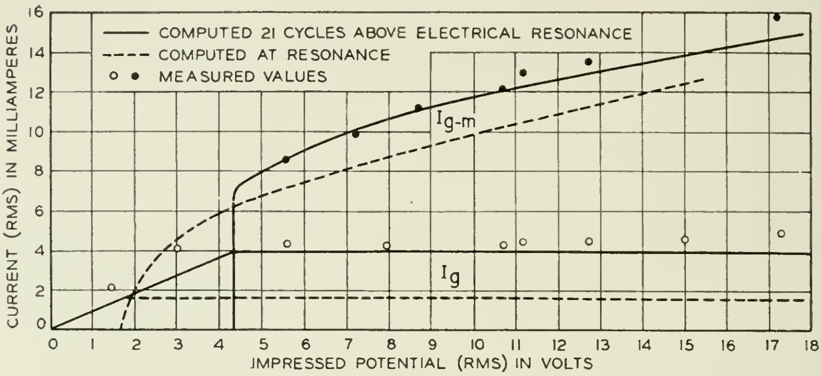


Fig. 2—Electrical current components.

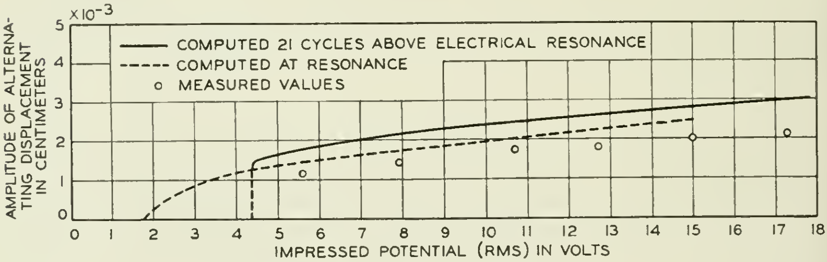


Fig. 3—Mechanical displacement.

While the frequency difference between them is only 20 cycles out of 30,000 it will be noted that, because of the critical character of the impedances, the results differ considerably in amplitudes of the components and in the threshold value.

The second case checks very closely as far as the electrical results are concerned and the mechanical results are of the same order of magnitude. The discrepancy can reasonably be laid to the inaccuracies in the indirect method of measuring the mechanical amplitude. More important is the verification of the outstanding properties predicted by the analysis. There is a threshold voltage above which

the new frequency components suddenly appear and rapidly build up to large amplitudes as the voltage is increased. The current of the impressed frequency, ω_g, remains practically constant and independent of voltage above the threshold.

There remains to be described the method by which the mechanical amplitude was obtained. The current voltage relation for a condenser is

$$V = \frac{1}{C}\int i\,dt.$$

The current through the condenser involves only three frequency components in this case, so it can be written in the form

$$i = A_g \cos(\omega_g t + \varphi_g) + A_d \cos[(\omega_g - \omega_m)t + \varphi_d)] + A_e \cos[(\omega_g - 2\omega_m)t + \varphi_e)]$$

and the capacity of a condenser, in e.s.u., is

$$C = \epsilon\frac{A}{S} = \frac{\epsilon A}{(S_0 + S_m \cos \omega_m t)} \text{ farads}$$

where

$$\epsilon = 8.85 \times 10^{-14} = \text{permittivity},$$

A = plate area in cm.^{2},
S_0 = constant, or average spacing cm.,
S_m = amplitude of mechanical displacement.

From these equations the amplitude of the mechanical vibration in terms of the electrical amplitudes can be determined. Neglecting phase angles the relation is

$$S_m = \frac{V_e - \dfrac{S_0 A_e}{\epsilon A(\omega_g - 2\omega_m)}}{\dfrac{A_d}{2\epsilon A(\omega_g - \omega_m)}}$$

V_e = amplitude voltage component of frequency $\omega_g - 2\omega_m$.

The neglect of the phase angles will make some inaccuracy in the results. They would be exact, in the above formula, were A_e negligibly small. The term involving A_e is a correction term necessitated by the incomplete suppression of that frequency component.

A New Type of Underground Telephone Wire

By D. A. QUARLES

A new type of telephone line is described in which a specially insulated *twin* wire is plowed into the soil. Problems of wire design, splicing and maintenance are discussed and transmission characteristics are given.

IN this day of multi-channel transmission on open-wire lines, lead-covered coaxial and multi-wire cables, and of radio and ultra-high-frequency transmission without lines at all, it behooves the development engineer concerned with line structures to be alert to advanced, even to radical, ideas. Rubber insulated telephone wire placed directly underground is a case in point.

The urge to put telephone lines underground is only a littler younger than the business itself. In large measure, this has been realized by installing lead-covered cables in underground duct systems. An alternative arrangement used more recently is spoken of as *buried cable.*[1] This is lead-covered cable, the sheath of which is protected from corrosion by successive layers of paper and jute flooded with asphalts. In addition, as a provision against mechanical injury or interference from outside electrical sources, a steel tape armoring is sometimes used. Where conditions have been favorable, the practice of burying suitably protected cables directly in the ground has been applied both to toll and to large and small exchange area cables and to one and two-pair entrance cables for underground service connections.

Because, with underground cables in conduits and with buried cables, the costs essentially limit their use to those cases where appearance is an important factor or where a considerable number of circuits can be grouped under the same sheath, these methods are not generally applicable to service on one or two circuit routes, such as those extending to remote subscribers, typical of rural distribution. Particularly in the interest of providing a lower cost type of plant and thereby making possible a more extensive development of service in rural communities, it appeared there would be a considerable incentive for the development of an inexpensive form of buried circuit. Such a circuit would obviously require that an economical means of installation be devised and even more important that the material used be serviceable for a

[1] C. W. Mier and B. D. Hull, *Bell Telephone Quarterly*, Volume 8 (October 1929).

long period under the severe moisture conditions to which it would be subjected in the ground.

Experience with the cable burial problem had led to the development of a cable laying plow, the neat operation of which in plowing cable into the ground at depths ranging up to thirty inches without trenching or backfilling in the ordinary sense has been described elsewhere.[2] The adaptation of this method to the burial of wire at an appropriate depth required that it be simplified so that it would be less expensive, and involved such considerations as the very much smaller size and tensile strength of the wire, its greater vulnerability to mechanical injury, the need for reducing and simplifying traction requirements, and the like.

On the point of serviceability, it remained for our research chemists first to develop a rubber compound that could be relied on to maintain suitable insulating properties over a period of years under the severe moisture conditions under ground.

With these fundamentals in hand, the development engineers undertook to study the mechanical and electrical problems involved and design a wire that would have appropriate transmission and handling characteristics. In addition, they had to devise methods of splicing the wire; to adapt plow equipment to its installation; to develop loading arrangements for use on the longer lengths; to study methods of tracing the path of the wire and locating faults, etc. In short, the job was to develop buried wire as a practicable plant instrumentality.

The Insulated Wire

The wire as actually developed employs 17-gauge annealed and tin-coated copper conductors, insulated in parallel twin construction with a special rubber compound designed to withstand long water immersion without serious deterioration of the electrical properties. The wire is adapted to a continuous process of extrusion and vulcanization.[3]

While the insulation of this wire is very resistant to water absorption, it is, in common with most high grade rubber insulating compounds, quite sensitive to sunlight so that it must be carefully guarded from any unnecessary exposure to direct rays of the sun and from any extended exposure to indirect rays.

Splicing

One of the principal problems in using a wire of this kind is that of splicing, as it is quite obvious that the splice must be essentially as

[2] C. W. Nystrom, *Telephony*, Volume 98 (June 21, 1930).
[3] S. E. Brillhart, *Mechanical Engineering*, Volume 54, pages 405–9 (1932).

resistant to water absorption as the wire itself. The splice actually developed (see Fig. 1) has two features of interest: the joints in the conductor proper and the method of patching the insulation. The

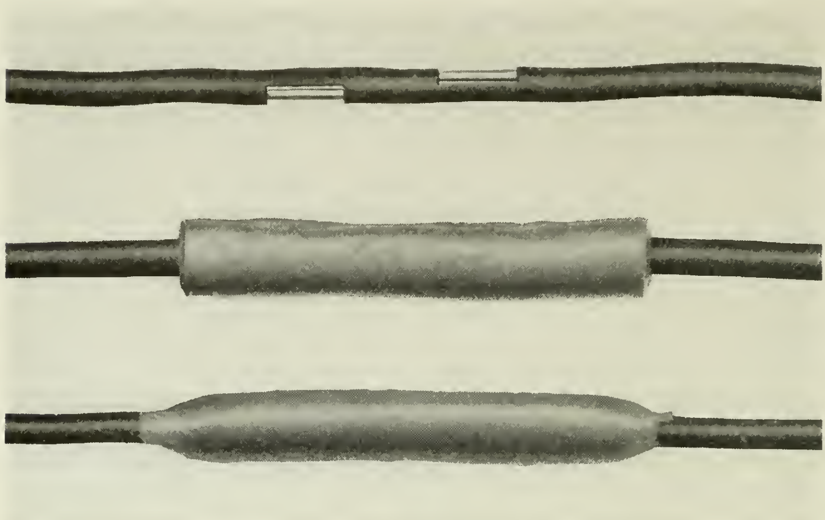


Fig. 1—Splicing buried wire. Top: pressed sleeve joints in conductors. Center: unvulcanized rubber pad in place. Bottom: after vulcanization.

conductor joint is made by pressing a cylindrical sleeve on the abutted ends of the wires to be joined, in this way producing a tight joint of high electrical efficiency and relatively immune to corrosion. The joints in the two wires are staggered and the whole encased in a pad of unvulcanized rubber which is pressed in place and vulcanized in an electrically heated mold, shown in Fig. 2. The vulcanizer is equipped with a thermostatic device to insure proper control of the temperature. This splice is intended for burial directly in the ground without other protection and tests indicate it to be the equivalent of the unspliced wire.

Electrical Properties and Loading

In cross-section, the insulated twin is an oval having a major diameter of about .33″ and a minor diameter of about .165″. The cross-section has been designed to give optimum electrical characteristics for the amounts of copper and rubber compound employed per unit length. The average mutual capacitance per thousand pair feet after seven days water immersion is about 0.022 mf. with an average loop resistance per thousand feet of about 10.2 ohms. While the trans-

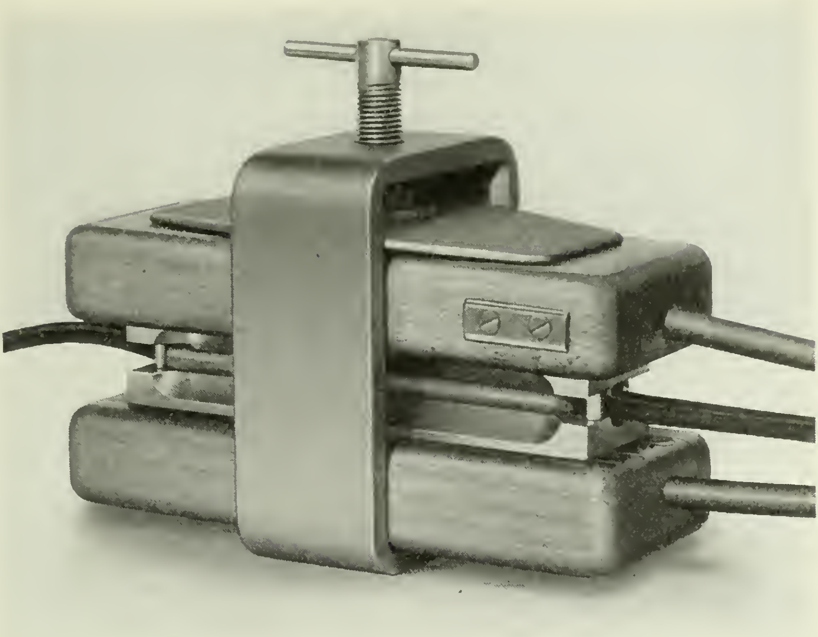


Fig. 2—Vulcanizer for buried wire splice. The buried wire is in the center and storage battery leads for heating vulcanizer are above and below on right.

mission requirements to be placed on a buried circuit will, of course, depend upon the facilities with which it is associated, it is expected that buried circuits up to about five miles in length will, in general, not require loading. Where loading is required, provision has been made for it in the form of a permalloy dust [4] core coil having 44 millihenries inductance which is individually potted with rubber insulated lead-out wires. It is intended to be spliced into the wire at 8,000-foot intervals and buried directly in the ground with the wire.

The potting arrangement for the buried wire coil has several features of interest. The loading coil is first potted in a small metal container which is vacuum impregnated with a moisture resistant compound. The lead-out wires from this container are then spliced to stub lengths of the buried wire, as shown in Fig. 3. This container is then placed in a larger sheet copper container, the rubber insulated wires being brought out through tubes soldered into the copper container and pressed down into intimate contact with the rubber insulation. The lead-out wires are taped for reinforcement at the outer ends of the

[4] G. W. Elmen, *Bell System Technical Journal,* Volume 15 (January 1936).

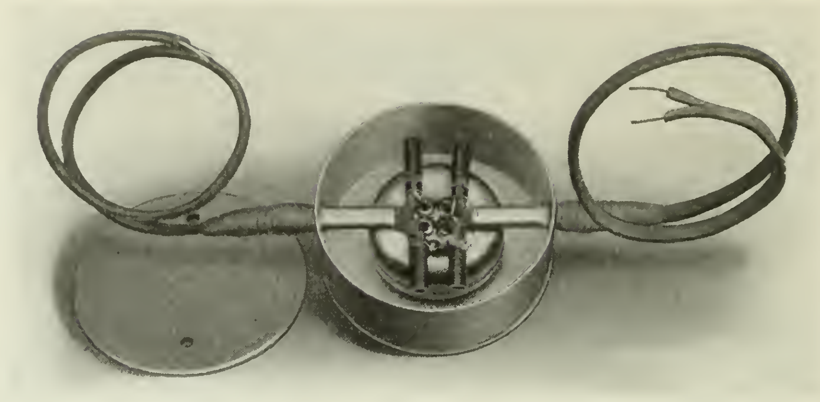


Fig. 3—Loading coil for buried wire before filling outer case, showing the splicing of the rubber covered stubs to the lead-out wires from the inner case.

tubes. This outer can is then filled with a moisture-proof compound and given a dip coating of moisture-proof enamel. The operation of splicing the loading coil into the line wire then involves making two line wire splices as above described.

The one-thousand-cycle attenuation of this 17-gauge buried wire after seven days water immersion and at 70° F. is about 1.1 db per mile for the non-loaded line and the corresponding attenuation of the loaded line is about .49 db per mile. The characteristic impedance of the non-loaded line at one thousand cycles and under the same conditions as above would be $275/\overline{40^\circ}$ and of the loaded line would be $525/\overline{8^\circ}$. The nominal cut-off frequency of the loaded circuit is 3600 cycles.

Layout of Buried Circuits

At the present time, the most promising use of buried wire in the telephone plant appears to be for rural distribution on routes requiring one or two pairs. These routes would commonly have a number of party line subscribers, each subscriber being bridged on the buried circuit. For the most part, it has been found preferable to follow the route of existing public roadways, laying the wire in the shoulder of the road. Installing the wire on right-of-way across private property is advantageous under some circumstances, however. At points where a service connection is to be made, there is the alternative of bridging a service lead across the through circuit or looping the circuit into the subscriber's house, the latter being preferred where the house is a short distance from the through route. Where a bridged connection is to be made, it has been found desirable to bring the wires up above

ground for binding post cross-connection so as to provide a test point rather than splicing the bridging wire in permanently and burying the splice in the ground. A small terminal has been provided for this purpose. The buried wires are brought up into the terminal housing through a galvanized iron pipe in order to provide protection of the rubber insulation from sunlight. The terminal is mounted on any convenient pole or post or on a short stub set for the purpose.

Fig. 4—Plowing-in two pairs of_buried wire along roadside.

As buried wire will, in general, be associated with exposed wire circuits, it is planned to provide the same type of electrical protection at subscribers' premises for buried circuits as for drops from open wire or exposed cable circuits. It is also planned to provide protection for buried wire at junctions with open-wire lines over one-half mile in length.

Plowing-in Operations

The success of buried wire is in considerable measure dependent upon the efficiency of the equipment provided for plowing it into the ground. This problem has therefore been studied carefully with a view to reducing the traction requirements to a minimum for the desired depth of placing, so as to permit the use of readily available tractive equipment. Experiments have indicated that in a given type of soil the tractive load on the plow increases approximately as the square of the depth of setting. The choice of depth is, of course, a

matter of judgment and is influenced somewhat by the local conditions. In general, it is felt that depths between 16″ and 20″ are adequate and that more shallow installations are justified only under special conditions.

The plow equipment that has been developed for this purpose is shown in operation in Fig. 4 and in more detail in Fig. 5. The plowshare is a vertical blade with a tube fastened to the back edge through

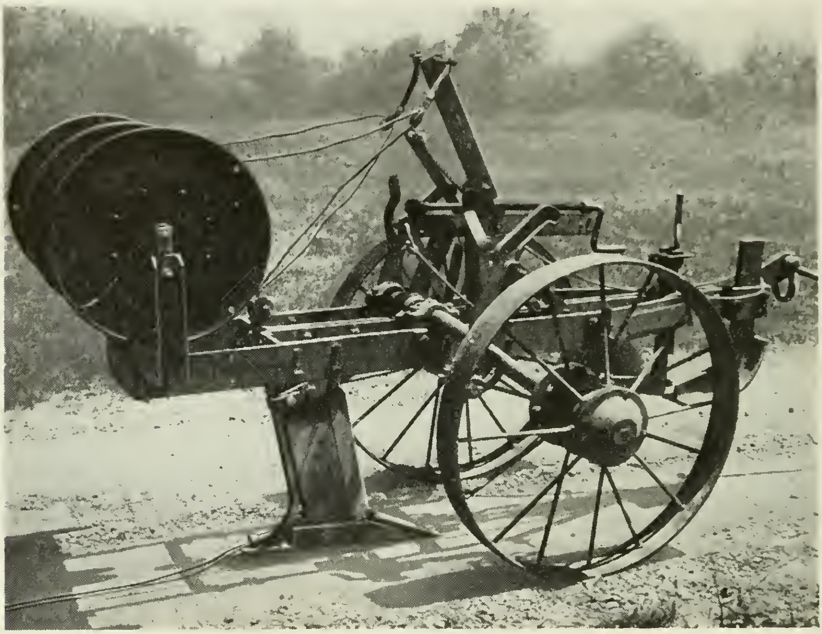


Fig. 5—Wire plow in elevated position, showing duct for wires at back of plow-share.

which one or two pairs of wires may be fed into the soil. The depth of the blade in the ground is readily adjustable to meet local conditions. It has been found that in fairly hard soil with a liberal supply of rock, an equivalent of a 40 or 50 hp caterpillar tractor is required to draw the plow. Across stretches of private right-of-way or at other locations where it is not convenient to use tractors or trucks in direct traction, however, alternative methods have been employed, such as using the winch line of a construction truck to pull the plow. The speed at which the plow may be operated is controlled largely by the number and character of obstacles encountered. Under favorable conditions, the plow may be operated at a speed of three or four miles an hour but a much lower average is to be expected under the less favorable conditions commonly found.

One consideration of some importance in installing wire of this type is the possibility of the insulation being crushed by boulders displaced by the plow, particularly where the trench with wire in place is to be rolled down or subjected to heavy traffic. This danger is, in fact, of such importance that buried wire of this type is probably not a serviceable form of construction through a terrain where nested boulders are frequently encountered.

While it is generally possible to plow across gravel highways, this method can not be used when hard-surface highways are encountered, and in such cases it becomes necessary either to use a pipe pushed under the roadway or to span the highway with open wire. Where conditions are such as to require routing the wire through or over culverts, across ditches, streams and the like, involving actual or potential exposure of the wire as by soil erosion, iron pipe or equivalent protection against mechanical injury and light will generally be required.

Interference

As in the case of other types of telephone circuits, the problem of avoiding noise and crosstalk must be considered. Where more than one pair is buried at a time, there is a crosstalk problem but experience has shown that the introduction of twists every few feet, either by twisting the wire in the process of laying or by having it pretwisted on the reels, is sufficient to reduce the crosstalk to low values.

Special care must be given the wire in manufacture to assure a good degree of balance between the capacitances of the two conductors to ground. This is important in order to avoid noise in the buried wire circuits when they are exposed to power circuits or when the connected open wire is exposed. Under severe exposure conditions, even with the best balance obtainable in manufacture, it may be necessary to resort to special balancing measures in the field to assure satisfactorily quiet circuits.

Maintenance Questions

The introduction of a new type of plant such as buried wire will naturally involve some new maintenance problems. Although records will, in general, be kept of buried wire routes, it will at times be desirable to have fairly precise methods of tracing the underground path. Experiments have indicated that this may be done with considerable precision by putting a tone current on the wire and following along the surface of the ground with an exploring coil device. The location of faults in buried wire also involves some problems which are different from those experienced with cable circuits but experiments have indicated that established methods may be adapted to this new use with an acceptable degree of precision.

Conclusion

As less than fifty miles of buried wire circuits of the type described have actually been installed and put into service, it is recognized that many problems may yet arise and that this type of plant should still be considered as in a trial stage. It is, for example, not known to what extent burrowing rodents such as gophers might cause difficulties. Soil erosion may also introduce problems not as yet clearly visualized. On the other hand, many wind, ice and tree interference troubles peculiar to open-wire construction, involving such things as broken insulators, broken poles, wires crossed or broken, etc., should be avoided by placing the wire under ground. Buried wire should also, in general, be free from lightning troubles when properly protected at junctions with open-wire lines. Considerations of this kind will be largely controlling in determining the eventual field of use for the buried type circuit. Present indications are that in any event many locations may be found where this type of construction will prove economical.

Effect of Electric Shock on the Heart *

By L. P. FERRIS, B. G. KING,† P. W. SPENCE and H. B. WILLIAMS †

AS a basis for the development of protective measures and practices, knowledge of the limits of dangerous electric shock is obviously important and this joint investigation at the College of Physicians and Surgeons of Columbia University was initiated in the hope of obtaining some of the needed data. In seeking a value of current which if exceeded would be dangerous to man, it is important to consider for different practical conditions the effects which are brought about as the current is increased. The threshold of sensation is reached at about one milliampere for a frequency of 60 cycles. Other investigators have found that at about 15 milliamperes from hand to hand the subject becomes unable to control the muscles subjected to stimulation.

Any currents that prevent voluntary control of the skeletal muscles are dangerous because their pathway through the body might include the respiratory muscles and stop breathing during the shock. If prolonged, asphyxial death would result, but the time required is a matter of minutes rather than seconds, so that opportunity may be afforded for action to release the victim. No serious or permanent after-effects are likely to appear merely from the cessation of respiration, provided it is not continued beyond the point where the victim can be resuscitated by artificial respiration.

Currents somewhat greater than those just necessary to stop respiration by action on the muscles may cause fatalities, even though the duration of such shocks is but a few seconds or less—far too short to be important from the standpoint of interruption of respiration and obviously too short to give any opportunity for rescue before the end of the shock. Death under such conditions is brought about by ventricular fibrillation, which is a disruption of normal heart action. This condition is an uncoordinated asynchronous contraction of the

* Digest of a paper published in the May 1936 issue of *Electrical Engineering* and presented at the A. I. E. E. Summer Convention, Pasadena, California, June 22–26, 1936.

† Drs. H. B. Williams and B. G. King, joint authors of this paper, are of the staff of the College of Physicians and Surgeons of Columbia University, New York.
Electrical facilities of the Bell Telephone Laboratories supplemented those of the Physiology Laboratories of the Medical Center where most of the experimental work was done.

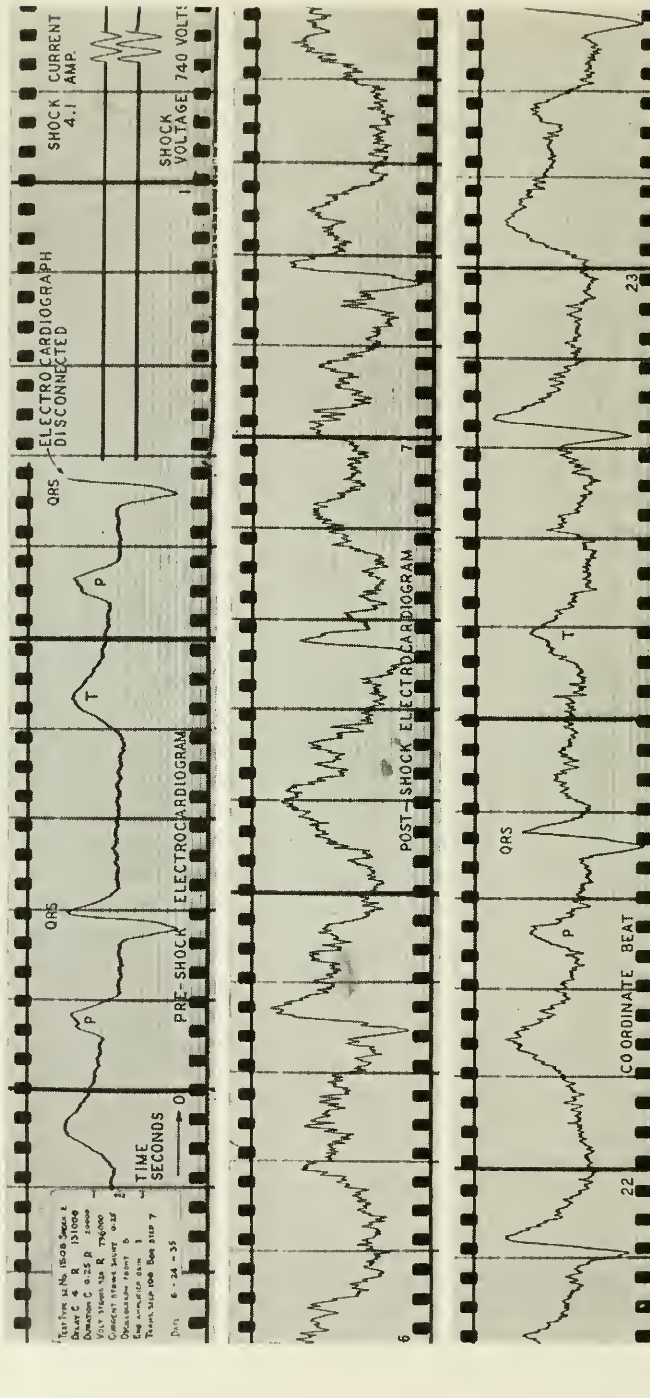


Fig. 1A—Typical records of 0.03-second shocks to sheep during insensitive phase of cardiac cycle. The electrodes were fastened to the right fore leg and left hind leg.

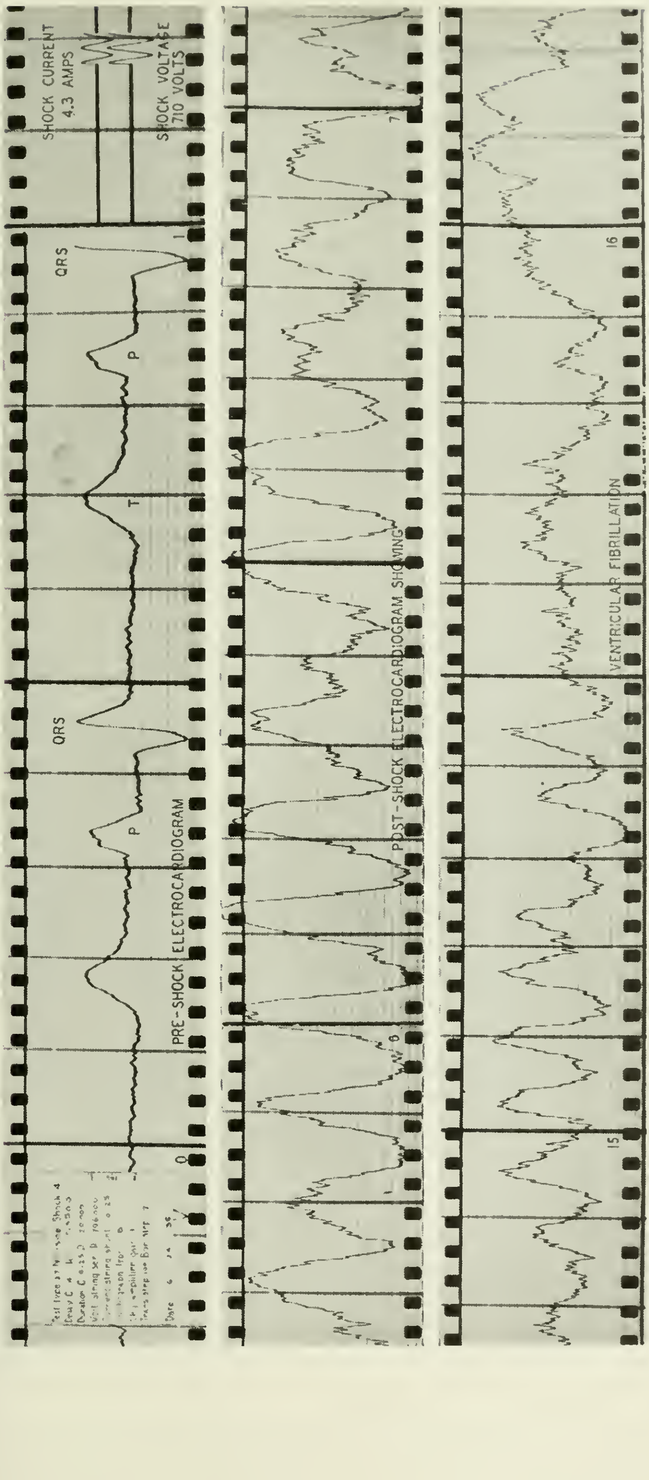


Fig. 1B—Typical records of 0.03-second shocks to sheep during sensitive phase of cardiac cycle.

ventricular muscle fibers in contrast to their normal coordinated and rhythmic contraction. It results from an abnormal stimulation rather than from damage to the heart. In the fibrillating condition, the heart seems to quiver rather than to beat; no heart sounds can be heard with a stethoscope; the pumping action of the heart ceases; failure of circulation results in an asphyxial death within a few minutes. The medical profession long has recognized that ventricular fibrillation once set up in man is unlikely to cease naturally before death. The value of current just under the threshold for ventricular fibrillation, therefore, may be taken as the maximum current to which man safely may be subjected, because regardless of rescue or after-treatment, death is liable to result from greater current.

This experimental investigation, therefore, was directed chiefly toward determining the minimum current that would initiate ventricular fibrillation and the variation of this threshold current with several factors which enter into the practical application of the results in the development of protective devices and measures. From the standpoints of both physiology and engineering, it was important to determine the influence on this threshold of:

1. Species and size of animal.
2. Path of current through the body (determined by points of contact).
3. Frequency of the current.
4. Time of occurrence of short shocks in relation to the cardiac cycle.
5. Duration of shock.

Thresholds were determined for seven species of animals: the guinea pig, rabbit, cat, dog, sheep, pig, and calf. Standard reference conditions included the use of 60-cycle alternating current for a duration of 3 seconds with electrodes on the right fore leg and left hind leg. These conditions typify those of many accidental shocks to man and are very dangerous from the point of view of ventricular fibrillation because the heart is almost directly in the current path.

Three significant records were made by an oscillograph for each shock. These are illustrated in Figs. 1A and 1B. They include electrocardiograms before and after shock and oscillograms of shock current and voltage. An electrocardiogram is a graphical record of the time variation of the voltage that is always associated with the action of the heart. The character of the variations in this voltage indicate certain facts as to the heart's condition, the electrocardiogram of a fibrillating heart being very different from that of a normal heart. The group of Fig. 1A shows a shock followed by coordinate beating. The group of Fig. 1B shows a shock which resulted in ventricular

fibrillation. The character of both post-shock electrocardiograms is somewhat masked by higher frequency currents resulting from skeletal muscle activity following the shock. However, the post-shock electrocardiogram of the group 1A reveals the same typical sequence of prominent deflections that appear before shock, while that of the group 1B shows an entirely different wave shape. The appearance of this typical fibrillating wave and the absence of heart sounds following shock were taken as conclusive evidence of ventricular fibrillation.

Threshold of Fibrillation

The threshold current increases roughly with both the heart weight and body weight of the different species of animals, although if the three smaller species be considered alone this relationship does not hold, their threshold currents being practically the same despite widely different body weights. Data from tests on a number of different species are summarized in Fig. 2.

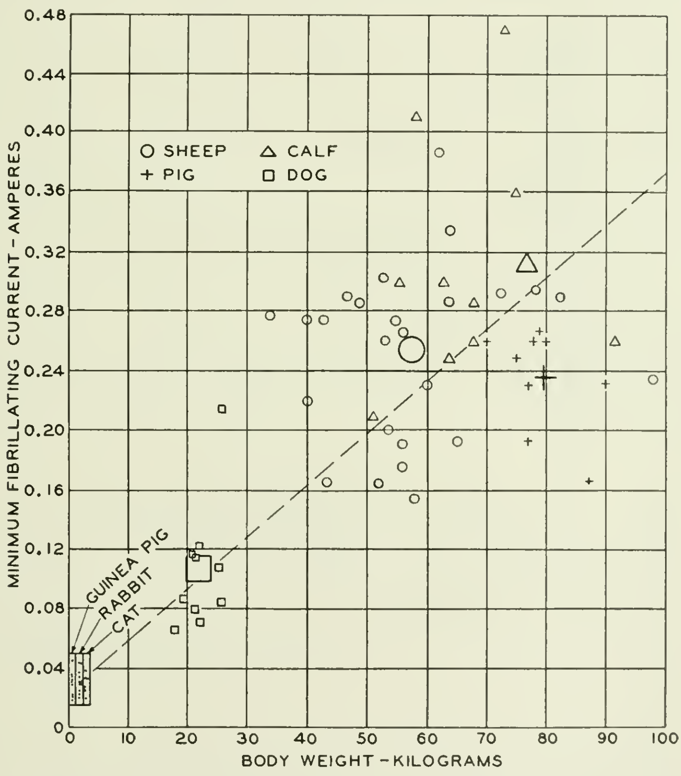


Fig. 2—Relation of minimum current causing ventricular fibrillation to body weight for different animal species. Shock duration 3 seconds. Frequency 60 cycles. Electrodes on right fore and left hind legs. (Averages shown by larger symbols.)

These results serve to indicate the probable threshold current for man under similar conditions. The average weight of an adult man is approximately 70 kilograms (154 lbs.) and his heart weight, 330 grams (.75 lbs.). The average threshold current for a body weight of 70 kilograms is 0.26 ampere and that corresponding to a heart weight of 330 grams is 0.29 ampere. Knowledge of such average currents is useful, but in the practical application of this information it is the lower limit of current causing ventricular fibrillation that must be taken into consideration. The thresholds differ widely for different individuals of the same species. The results on the whole indicate for man that currents in excess of 0.1 ampere at 60 cycles from hand to foot would be dangerous for shock durations of three seconds or more.

Effect of Frequency

To determine the effect of changing the frequency other tests were made on sheep with 25-cycle current and direct current, the general

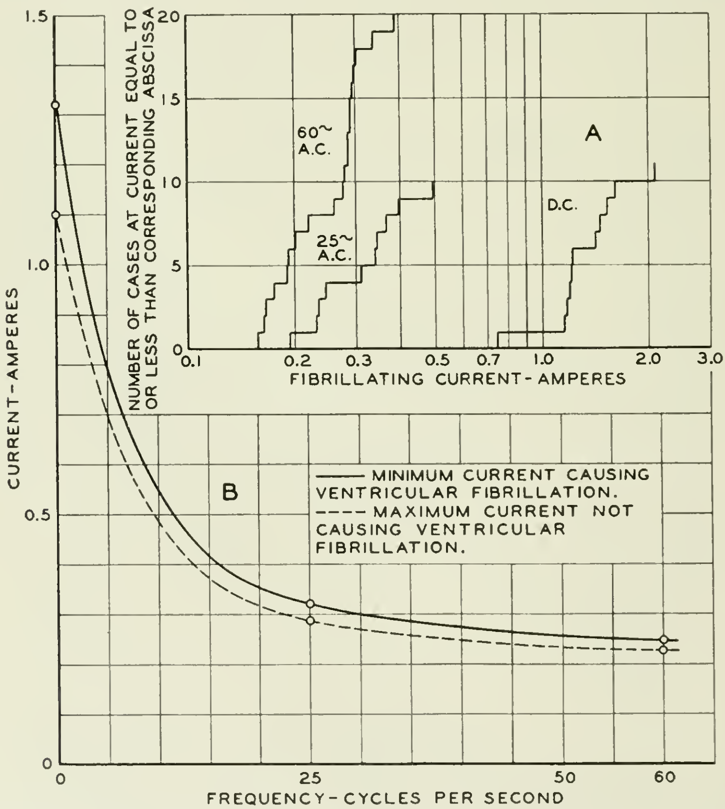


Fig. 3—Effect of frequency on threshold current causing ventricular fibrillation in sheep. Shock duration 3 seconds. Electrodes on right fore and left hind legs.

conditions remaining the same as for the 60-cycle current. The results of these tests are illustrated graphically in Fig. 3.

Susceptibility of Heart in Different Phases of Cardiac Cycle

Physiologists have established that the heart is responsive to moderate electrical stimuli during the period of relaxation (diastole) whereas such stimuli during the period of contraction (systole) do not elicit further response. During early systole (*QRS* to beginning of *T*, Fig. 1) the heart muscle is totally unresponsive, while during diastole (end of *T* to beginning of *QRS*) the ventricular muscles are responsive to stimuli. At the time of the *T* wave of the electrocardiogram, contraction starts to disintegrate and parts of the ventricular muscle will respond differently to electrical stimuli. In view of these facts and some erratic responses to shocks of three to four cycles of 60-cycle current, differences were expected in the response of the heart to very short shocks during different phases of its cycle. Special apparatus was developed for applying short shocks at predetermined parts of the cardiac cycle. The results of 913 shocks of 0.03-second duration on 132 sheep are plotted in Fig. 4, to show the position of the mid-point of each shock in the cardiac cycle, approximate shock current and whether or not fibrillation occurred. None of the shocks occurring during the period of complete contraction or complete relaxation of the heart caused ventricular fibrillation, this result appearing only for shocks falling during the period of diminishing contraction.

Of 370 shocks of 0.12 second duration applied to 38 sheep, only one shock definitely outside the partial refractory phase resulted in ventricular fibrillation. This shock began at a point in the electrocardiogram between *P* and *Q* waves at which time the ventricles are completely relaxed and resting.

Effect of Duration

In the determination of thresholds for shocks of very short durations, a third or less of the duration of one heart beat, the time of occurrence of the shocks was regulated so as always to involve the partial refractory phase, corresponding to the appearance of the *T* wave. The thresholds were found in the same way as for 3-second shocks, by applying successive shocks at intervals of 5 minutes, each at increased current until ventricular fibrillation resulted.

The effect of duration on the threshold is illustrated in Fig. 5. It is believed that the current required to initiate fibrillation would increase markedly as the duration is decreased below 0.03 second. It

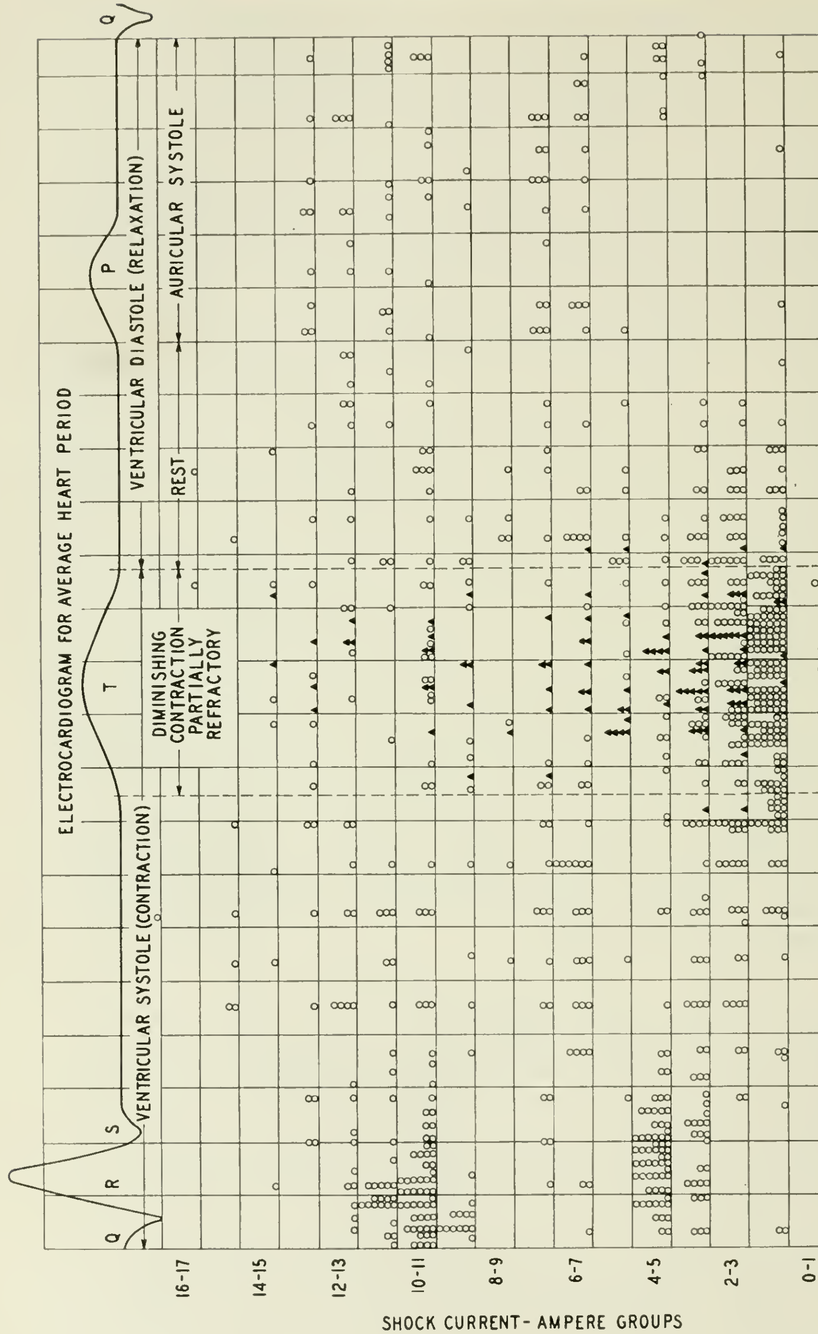


Fig. 4—Distribution in cardiac cycle and results of 913 shocks of 0.03-second duration to 132 sheep. 60 cycles. Electrodes on right fore and left hind legs. Triangle indicates fibrillation. Circle indicates coordinate beat.

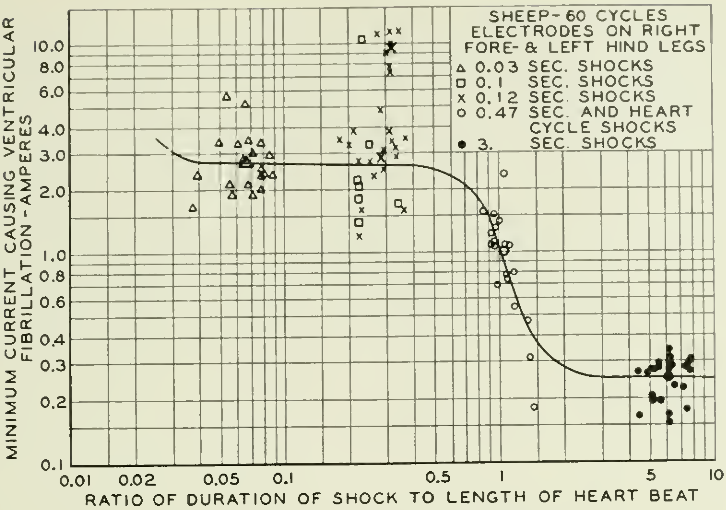


Fig. 5—Effect of duration on threshold.

is, however, not believed that increasing the duration beyond 3 seconds would reduce the average threshold current appreciably below 0.25 ampere.

Ventricular fibrillation has been found to be the only serious cardiac effect of the currents applied in these tests; however, temporary disruptions of normal cardiac activity frequently can be observed. A most common effect of electric shock is a change in heart rate. Electrocardiograms after shock frequently indicate disturbances of conduction in the heart. Premature heart beats (extrasystoles) and fibrillation of the auricles have also been observed.

The persistence of any of these conditions for more than a few minutes is rare. There is no evidence of any cardiac abnormalities or the presence of cardiac damage in electrocardiograms taken at intervals up to two months following shocks which did not immediately cause death.

Effect of High Currents

There was evidence from the work of Prevost and Battelli and some of the early results of this investigation that the stimulus of a powerful current would be less liable to cause fibrillation than a current moderately above the threshold. To test such evidence repeated short shocks of 23 to 26 amperes were given to a group of sheep in the sensitive phase of their cardiac cycle. Ten survived 5 shocks each without fibrillation, while an eleventh fibrillated on the initial shock. Each of the 10 surviving sheep was given additional similar shocks,

except that the current was reduced to between 4 and 5 amperes. Five developed ventricular fibrillation on the first shock, and 3 on the second shock. A single animal survived 5 shocks, and another animal 2, without fibrillating. A comparison and combination of these results and those previously obtained definitely establish that the susceptibility of the heart to ventricular fibrillation becomes very much less when the current is increased to about 25 amperes, 10 times the average threshold value for shocks of short duration. This variation of susceptibility to fibrillation is shown graphically in Fig. 6.

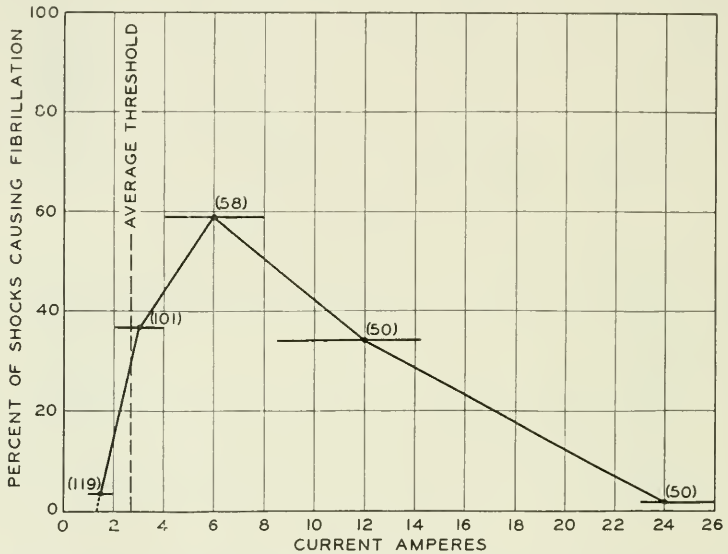


Fig. 6—Effect of current on susceptibility of sheep hearts to ventricular fibrillation. Shocks of 0.03 second, 60 cycles, applied in partial refractory period of cardiac cycle. Number of shocks and current spread indicated for each point on curve.

To determine whether such a decrease in susceptibility to fibrillation would occur also for shocks of about 3 seconds duration if the current were increased about 15 times the average threshold, 5 sheep were subjected to 4 ampere shocks of 3 seconds duration. In all cases ventricular fibrillation resulted on the initial shock, indicating either that at this duration no such decrease in susceptibility takes place with increase of current, or that 4 amperes is not a sufficiently high current to bring it about. The apparatus was not capable of giving much higher currents for this duration.

except that the current was reduced to between 4 and 5 amperes. Five developed ventricular fibrillation on the first shock, and 3 on the second shock. A single animal survived 5 shocks, and another animal 2, without fibrillating. A comparison and combination of these results and those previously obtained definitely establish that the susceptibility of the heart to ventricular fibrillation becomes very much less when the current is increased to about 25 amperes, 10 times the average threshold value for shocks of short duration. This variation of susceptibility to fibrillation is shown graphically in Fig. 6.

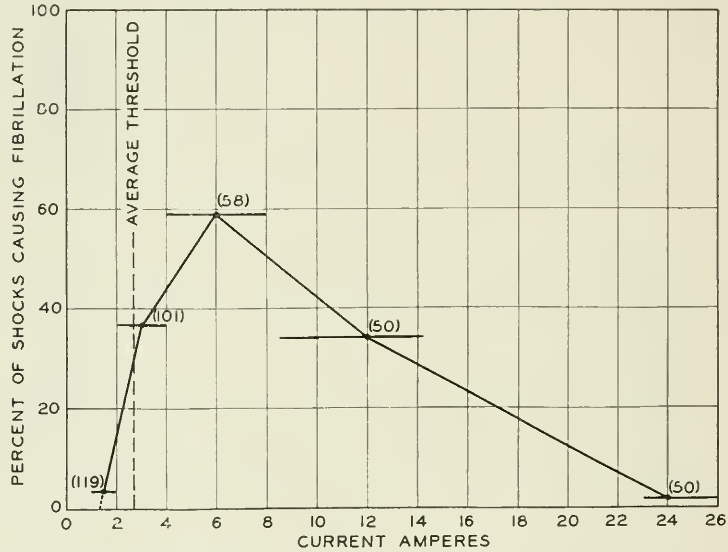


Fig. 6—Effect of current on susceptibility of sheep hearts to ventricular fibrillation. Shocks of 0.03 second, 60 cycles, applied in partial refractory period of cardiac cycle. Number of shocks and current spread indicated for each point on curve.

To determine whether such a decrease in susceptibility to fibrillation would occur also for shocks of about 3 seconds duration if the current were increased about 15 times the average threshold, 5 sheep were subjected to 4 ampere shocks of 3 seconds duration. In all cases ventricular fibrillation resulted on the initial shock, indicating either that at this duration no such decrease in susceptibility takes place with increase of current, or that 4 amperes is not a sufficiently high current to bring it about. The apparatus was not capable of giving much higher currents for this duration.

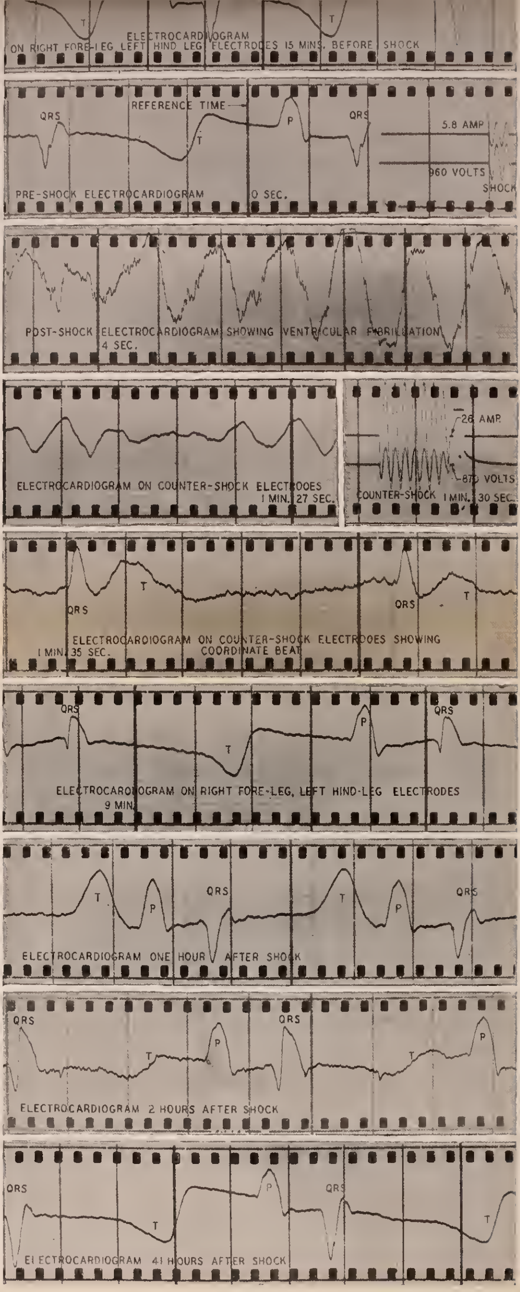


Fig. 7—Record of restoration of heart from ventricular fibrillation by a counter-shock.

Recovery of Heart from Ventricular Fibrillation

Recovery of the heart from ventricular fibrillation by the application of a short intense shock was reported first by Prevost and Battelli in 1899. Kouwenhoven, Langworthy and Hooker also have reported the recovery of dogs from fibrillation by the application of what has recently been termed a "counter-shock." The opportunity to experiment on recovery from ventricular fibrillation with large animals arose with the use of sheep and other large species in 1932. Currents up to 25 or 30 amperes were applied through large electrodes placed near the heart.

After some preliminary tests all counter-shocks were given with electrodes outside the skin on both sides of the chest so as to include the heart between them. Such counter-shocks were found to be effective in recovering coordinate heart action in about 60 per cent of the cases.

Figure 7 is an electrocardiographic record of heart action before and after a shock that caused fibrillation and at different stages after the application of a counter-shock that arrested the fibrillation and allowed the heart to resume its coordinate beating. The current and voltage of the fibrillating shock and the counter-shock are shown also to the same scale for comparison. The fibrillating shock of 6 amperes and 0.03 second duration occurred during the sensitive phase of the cardiac cycle. The counter-shock which followed 1½ minutes later was of 26 amperes for 0.1 second duration. Times marked on the different sections are referred to the beginning of the record. It may be observed that the last electrocardiogram is practically identical with the pre-shock electrocardiogram. Many sheep have been observed for periods of months after recovery from fibrillation, with no evidence of abnormalities. Several have given birth to normal lambs, and in many instances the recovered sheep have been used in subsequent tests and again recovered.

The possibilities of counter-shock have not been fully explored to determine the optimum conditions for its application, particularly as regards magnitude and character of current, its duration, and points of application. In regard to the latter, however, it would seem that some short path embracing the heart would be best. Any technique of recovery of the heart must be applied promptly so as not to permit deterioration of the brain which might result in impairing the competency of the victim if recovered. While the time limit would depend on many factors, it is a matter of minutes rather than seconds. The prompt application of artificial respiration ventilates the lungs and is believed also to maintain a small circulation of blood, sufficient to delay

degeneration in the brain. This is of fundamental importance in the development of practical recovery methods. Wiggers recently has pointed out that maintenance of coronary circulation is essential for recovery from fibrillation.

Were suitable arrangements and methods developed for the practical application of counter-shocks, such shocks might be applied mistakenly to a victim whose heart was not in ventricular fibrillation. To determine whether in such an event ventricular fibrillation would be caused, 5 "counter-shocks" of about 25 amperes for about 0.06 second were applied to each of 9 sheep whose hearts were beating normally. Only 3 of the 45 shocks applied caused fibrillation, and in every such case recovery was obtained by the immediate application of another similar shock. This experiment was performed in 1932 prior to the development of apparatus for the controlled placement of short shocks, so that the shocks naturally fell at random. It was also prior to the determination that such high-current shocks of short duration were not likely to cause fibrillation even when the shock occurred during the sensitive phase of the heart cycle. In the light of the subsequent experiments, it is evident that the liability of causing ventricular fibrillation by randomly placed short shocks at the high currents employed in counter-shock is small.

Summary of Results and Conclusions

1. Current rather than voltage is the proper criterion of shock intensity.

2. The stimulating effect of current through the heart can derange its action, causing ventricular fibrillation without damage to the cardiac tissues but resulting in death within a few minutes, unless the fibrillation is arrested.

3. A current just below the threshold for ventricular fibrillation is the maximum to which man safely may be subjected. Based upon numerous tests on animals of several species comparable in size with man, this maximum current is about 0.1 ampere at 60 cycles for a duration of one second or more and a pathway between an arm and a leg.

4. The threshold fibrillating current is affected by:

a. Species and Size of Animal. Among the different species the threshold current increases roughly with both body weight and heart weight.

b. Current Pathway. Pathways from arm to leg, across the chest, chest to arm, and head to leg may be expected to give about the same threshold current. The pathway between the arms would be expected

to give a somewhat higher threshold current. For the pathway from leg to leg, the proportion of current reaching the region of the heart is so small that fibrillation is not liable to result, even at currents of 15 amperes or more, although such currents probably would burn the victim unless the contacts were good and the shock of short duration.

c. Frequency. For shocks of one second or more in duration, the 25-cycle threshold current is about 25 per cent higher than the 60-cycle value, and the d-c. threshold current 5 times the 60-cycle value. For shock durations of a small fraction of a second this relation probably does not hold, all thresholds being expected to approach one another.

d. Time of Occurrence of Short Shocks in Relation to Cardiac Cycle. The heart is most sensitive to fibrillation for shocks occurring during the partial refractory phase of its cycle, which is about 20 per cent of the whole and which occurs simultaneously with the T wave of the electrocardiogram. With shocks of a duration of about 0.1 second or less, it is practically impossible to produce ventricular fibrillation, unless such shocks coincide in part at least with this sensitive phase of the cardiac cycle. The middle of the partial refractory phase is more sensitive than its beginning or end.

e. Duration of Shock. The threshold current varies inversely with shock duration but not uniformly, being most sensitive to change as the duration approaches the duration of one heart beat. Within the sensitive phase of the heart cycle the threshold fibrillating current for shock durations of about 0.1 second or less is 10 or more times the threshold for durations of one second or more. Shocks $\frac{1}{3}$ or more of the heart cycle in duration may cause ventricular fibrillation, even though they would not extend into the sensitive phase of the cycle if the heart continued its normal beat after the initiation of the shock. The reason for this is probably the initiation of a premature heart beat which brings about a premature sensitive phase prior to the end of the shock.

5. Successive shocks have no cumulative effect on the susceptibility of the heart to fibrillation.

6. The susceptibility of the heart to fibrillation by short shocks increases with current up to several times the threshold, then diminishes, becoming very small at currents of the order of 25 amperes through the body in the vicinity of the heart. However, other serious injury may be expected from such currents when brought about by accidental contacts.

7. Fibrillation produced by an electric shock will in the majority of cases be arrested by a subsequent electric shock of high intensity and short duration through the heart, allowing the resumption of coordinate beating with no permanent damage.

With about 60 per cent success in recovering animals comparable in size to man from ventricular fibrillation by the application of a rather arbitrarily chosen counter-shock, further study is desirable to develop the optimum conditions and practical apparatus for utilizing this method in accident cases. The use of a simple electrocardiograph appears desirable since ventricular fibrillation cannot be recognized positively by stethoscopic examination.

Should a counter-shock be applied mistakenly to a coordinately beating heart, the liability of its causing fibrillation is small and, should this occur, another counter-shock probably will arrest the fibrillation and bring back coordinate heart action.

To be successful, a counter-shock must be administered promptly after the fibrillating shock, probably within a few minutes.

The use of a counter-shock does not in any way lessen the need for maintaining respiration, by artificial means if necessary. In fact, the administration of artificial respiration even in the interval before application of a counter-shock is highly advisable, not only for respiration itself, but because of the accompanying slight circulation which will assist in the nutrition of the heart and delay degeneration of the brain.

A New High-Efficiency Power Amplifier for Modulated Waves *

By W. H. DOHERTY

THE use of increasingly higher power levels in broadcasting in the last few years has attached new importance to the matter of more efficient operation of the high-power stages in radio transmitters. The resulting reductions in cost of power, size of high-voltage transformers and rectifier, and water cooling requirements, are of particular importance in transmitters having outputs of 50 kilowatts or more.

The linear radio-frequency power amplifier, in the form in which it has been used extensively in broadcast transmission, may not be operated at a plate efficiency higher than about 33 per cent, if it is to supply the peak power required for amplifying a completely modulated wave. With this efficiency the d-c power input to a 50-kilowatt amplifier, for example, is 150 kilowatts, of which 100 kilowatts must be dissipated at the anodes of the water-cooled tubes.

This inherent weakness of the conventional linear amplifier has occasioned the development of certain other systems of amplification or modulation which permit a more economical use of power. Of these the high-level Class B modulation system [1] and the ingenious "outphasing modulation" scheme of Chireix [2] are most worthy of note.

A new form of linear power amplifier has been developed which removes the limitation of low efficiency inherent in the conventional circuit, permitting efficiencies of 60 to 65 per cent to be realized, while retaining the principal advantages associated with low-level modulation systems and linear amplifiers, namely, the absence of any high-power audio equipment, the ease of adding linear amplifiers to existing equipment to increase its power output, and the adaptability of such systems to types of transmission other than the carrier-and-double-side-band transmission most common at present.

Linear radio-frequency power amplifiers are ordinarily biased

* Digest of a paper presented before the Annual Convention of the Institute of Radio Engineers, May 11–13, 1936, at Cleveland, Ohio, and to be published later in the *Proceedings*.

[1] Chambers, Jones, Fyler, Williamson, Leach, and Hutcheson, "The WLW 500-Kilowatt Broadcast Transmitter," *Proc. I. R. E.*, Vol. 22, p. 1151, October, 1934.

[2] Chireix, "High Power Outphasing Modulation," *Proc. I. R. E.*, Vol. 23, p. 1370, November, 1935.

nearly to the cut-off point so that the plate-current pulse width is approximately a half cycle. Under these conditions the plate efficiency is proportional to the amplitude of the radio-frequency plate voltage. It is possible to obtain large outputs from tubes with radio-frequency plate voltage amplitudes of 0.85 to 0.9 of the applied d-c potential, i.e., with the plate voltage swinging down to a minimum value as low as 10 to 15 per cent of the d-c potential. The corresponding plate efficiency for the tube and its tuned circuit is approximately 67 per cent. This condition, however, prevails only at the peak output of the amplifier, and since the amplitude of the plate voltage wave, in a transmitter capable of 100 per cent modulation, is only half as great for the unmodulated condition as for the peaks of modulation, the efficiency with zero modulation in the conventional amplifier does not exceed half this peak value, or about 33 per cent. Even during complete modulation the effective efficiency over the whole audio cycle is only 50 per cent, and for the average percentage modulation of broadcast programs the all-day efficiency is only slightly greater than the efficiency for unmodulated carrier.

In order to improve this situation it is necessary to devise a system in which the amplitude of the alternating plate voltage wave is high for the unmodulated condition, and in which the increased output required for the positive swings of modulation is obtained in some other manner than by an increase in this voltage.

A simple and fundamental means is available for achieving this result. One embodiment of the scheme is illustrated in Fig. 1. Each

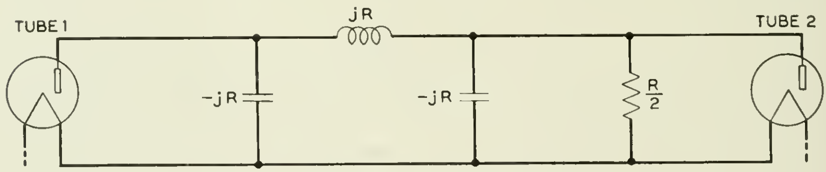


Fig. 1—Form of high-efficiency circuit.

of the two tubes shown in this figure is designed to deliver a peak power of E^2/R watts into an impedance of R ohms. The total peak output of the two tubes being $2E^2/R$ watts, the tubes are suitable for use in an amplifier whose carrier output is one-fourth of this value, or $E^2/2R$ watts. If the tubes were to be connected in parallel in a conventional amplifier circuit the load impedance used would be $R/2$ ohms, and each tube would work effectively into R ohms by virtue of the presence of the other tube. The same load impedance $R/2$ is used in the new circuit, but between the load and one of the tubes

(Tube 1) there is interposed a network which simulates a quarter-wave transmission line at the frequency at which the amplifier is operated.

It is a well-known property of quarter-wave transmission lines and their equivalent networks that their input impedance is inversely proportional to the terminating impedance. The network of Fig. 1, in particular, presents to Tube 1 an impedance of R ohms when its effective terminating impedance is also R ohms, that is, when half of the power in the load is being furnished by Tube 2; but should Tube 2 be removed from the circuit, or prevented from contributing to the output, the terminating impedance of the network would be reduced to $R/2$ ohms, with a consequent increase in the impedance presented to Tube 1 from R to $2R$ ohms. Under this condition Tube 1 could deliver the carrier power $E^2/2R$ at its maximum alternating plate voltage E and consequently at high efficiency.

The operation of the amplifier over the modulation cycle is as follows: The grids of both tubes are excited by the modulated output of the preceding amplifier stage, but for all instantaneous outputs from zero up to the carrier level Tube 2 is prevented by a high grid bias from contributing to the output, and the power is obtained entirely from Tube 1, which is working into $2R$ ohms, twice the impedance into which it is to work when delivering its peak output. In consequence, the radio-frequency plate voltage on this tube at the carrier output is nearly as high as is permissible and the efficiency is correspondingly high. Beyond this point the dynamic characteristic of Tube 1, unassisted, would flatten off very quickly because the plate voltage swing could not be appreciably increased. Tube 2, however, is permitted to come into play as the instantaneous excitation increases beyond the carrier point. In coming into play Tube 2 not only delivers power of itself, but through the action of the impedance-inverting network causes an effective lowering of the impedance into which Tube 1 works, so that Tube 1 may increase its power output without increasing its plate voltage swing, which was already a maximum at the carrier point. At the peak of a 100-per-cent modulated wave each tube is working for an instant into the impedance R most favorable to large output and delivering E^2/R watts, twice the carrier power, so that the total instantaneous output is the required value of four times the carrier power. Thus the required tube capacity is the same as in a conventional linear power amplifier.

Since it is usually desirable in power amplifiers to provide low-impedance paths for the harmonic components of the radio-frequency plate current wave, the reactive elements designated $-jR$ in a practical circuit ordinarily consist of a considerably larger capacity shunted by

a coil, and in the tuning process either the coil or the condenser is adjusted so that the impedance of the combination is the required value of $-jR$ ohms. The effective shunt load $R/2$ is then usually obtained by coupling the radiating system of the transmitter to the necessary extent into the parallel circuit associated with Tube 2.

The presence of a quarter-wave network in the output circuit of the amplifier causes the plate potentials on the two tubes to be 90 degrees apart in phase. This requires that the voltages impressed on the two grids be 90 degrees apart in order that each may be opposite in phase to the related plate potential, as is necessary in any power amplifier. In addition to this phase requirement, there arises from the variation in load impedance for Tube 1 the requirement that the excitation on this tube shall rise considerably less than 100 per cent on the positive peaks of modulation. Without some limiting action on this excitation the grid current in Tube 1 would be excessive and would result in a diminution of its output at modulation peaks.

Both of these requirements concerning the input to the amplifier are satisfied by the use of the input circuit shown in Fig. 2. With the

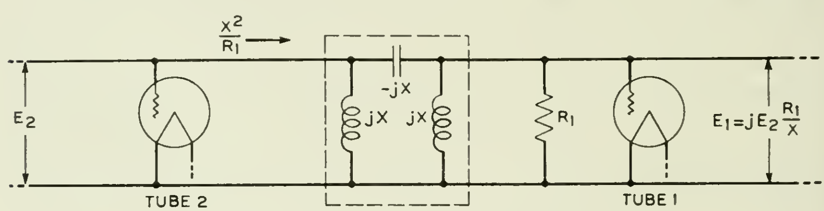


Fig. 2—Input circuit for a high-efficiency amplifier.

output of the previous stage applied directly to the grid of Tube 2, the resulting excitation E_1 on Tube 1 is proportional to the terminating resistance R_1 of the quarter-wave network. With a suitable value of R_1 the input conductance of Tube 1 arising from the flow of grid current at high excitations causes an effective lowering of R_1 which gives the desired limiting action on the excitation. At the same time the input impedance X^2/R_1 of the quarter-wave network is increased, compensating to a large extent for the shunting effect of the grid current in Tube 2, so that the previous stage is assisted in maintaining the proper excitation on the amplifier.

In a preliminary study of the behavior of an amplifier under these new conditions of operation, the results shown in Fig. 3 were obtained with a pair of small tubes. The radio-frequency plate potential of Tube 2 is the potential across the load circuit and is required to be linear with excitation. The short dotted portion halfway up on this

characteristic shows the curvature that would be obtained if Tube 2 were not allowed to come into action. With proper adjustment of the bias and relative excitation on Tube 2 this effect is eliminated and the characteristic continues to rise up to the desired peak amplitude.

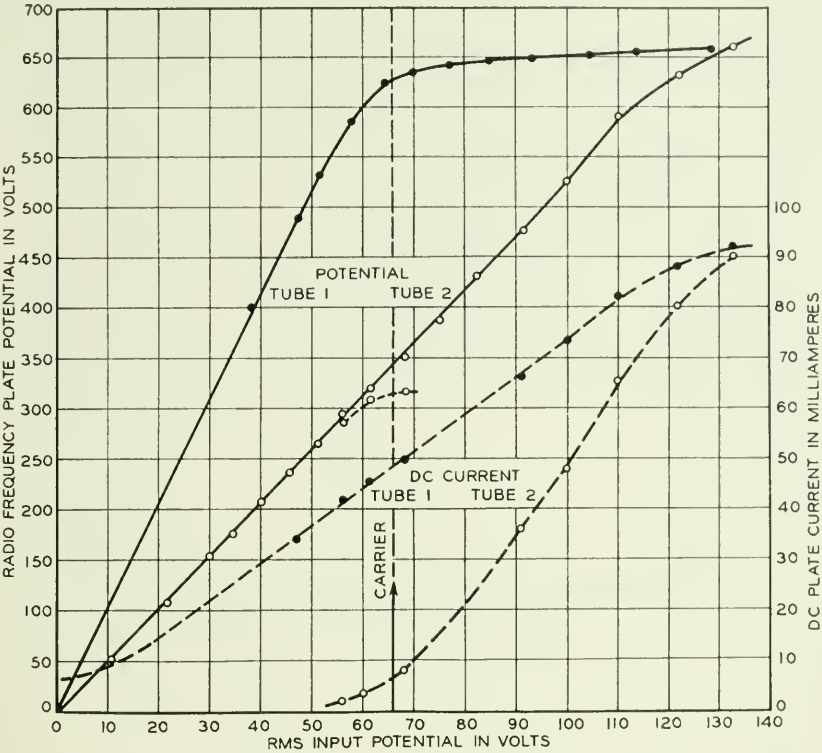


Fig. 3—Dynamic characteristics of an experimental low-power high-efficiency amplifier.

The radio-frequency plate voltage of Tube 1 is seen to be twice that of Tube 2 up to the point where curvature begins, and then to increase only slightly between the carrier output and peak output. The plate current of Tube 2 commences just before the carrier point is reached and rises twice as rapidly as the plate current of Tube 1. The equality of plate currents and radio-frequency plate potentials on the two tubes at the peak of modulation indicates that the tubes are contributing about equally to the instantaneous output at this point.

The high radio-frequency plate potential of Tube 1 at the carrier amplitude results in an efficiency of 63 per cent, and by integrating the

d-c plate currents of Fig. 3 over a complete cycle of modulation it is found that the effective average efficiency at 100 per cent modulation is also 63 per cent. The d-c plate current of the amplifier therefore rises 50 per cent with full modulation, as does the output power.

The necessity for careful adjustment of the relative excitation and bias on Tube 2, to obtain a linear characteristic in the amplifier, is eliminated when the feedback principle[3] due to Black is employed. Negative feedback may be used in radio transmitters at either radio frequency or audio frequency. The resulting improvements in linearity are useful in noise reduction as well as in distortion correction.

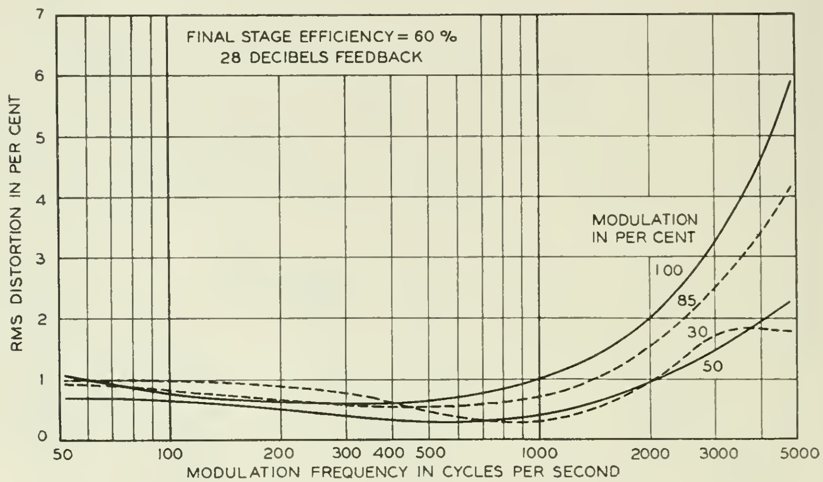


Fig. 4—Distortion measurements on a 50-kilowatt transmitter.

Figure 4 gives the results of distortion measurements on a complete 50-kilowatt transmitter built in the Laboratories and operating at a plate circuit efficiency for the final stage of 60 per cent. The use of 28 db of audio-frequency feedback, besides permitting alternating-current filament heating for all of the tubes, resulted in a distortion level less than 1 per cent at any frequency between 50 and 1000 cycles. At the higher audio frequencies the feedback is less effective because of the cumulative phase shifts in the various stages. The percentage modulation actually occurring in a broadcast program at these high frequencies, however, is so small that the distortion measured at high percentages of modulation is not of practical significance. The test, moreover, was made at the low-frequency end of the broadcast

[3] H. S. Black, "Stabilized Feedback Amplifiers," *Electrical Engineering*, January, 1934; *Bell Sys. Tech. Jour.*, January, 1934.

spectrum, where the effect is most pronounced because of the smaller band width.

The power required by this transmitter, including all auxiliary equipment, was 135 kilowatts with normal program modulation, as compared with approximately 230 kilowatts required in the usual 50-kilowatt installation.

High-efficiency operation, in addition to affording a large saving in the plate power supply, reduces the plate dissipation by a factor of three or four, with a resulting economy in the cooling system and an improvement in tube life.

The absence of any such requisites as the complicated driving stages of the Chireix system or the large audio equipment involved in high-level modulation gives the new circuit an important advantage over other high-efficiency systems in cost of apparatus and simplicity of design. The new amplifier, moreover, is operated at a plate voltage consistent with safety to the tubes, and is therefore not subject to the operating difficulties encountered at the high peak plate voltages required in high-level modulation.

Abstracts of Technical Articles from Bell System Sources.

A Review of Radio Communication in the Mobile Services.[1] CLIFFORD N. ANDERSON. Developments in radio communication in the mobile services during 1935 have been largely in the nature of gradual improvement of existing equipments and services.

In the marine field, the safety-of-life aspect is assuming increased importance. Rearrangements have been made of the frequencies and the schedules of radio beacons to avoid interference thereby making the system more effective. The improvement of radio compasses, regulations regarding motor lifeboat equipment and public address alarm systems, requirements for radio auto alarms and experimentation with collision prevention equipment are other items on which progress has been made the past year. The development in marine radiotelegraphy has been chiefly along the lines of greater application of the high frequencies. Directional antennas at the shore receiving stations have mitigated the effects of interference. Facsimile transmission of weather maps and press is being tried out. Improvements have been made in radio telephone equipments of various powers and frequency ranges for various types of marine service. A system utilizing ultra-high frequencies was put into operation at Philadelphia during the year. Three commercial stations are in operation in the two-megacycle range, the one at Seattle having been opened this year.

Radio is an important factor in the operation of modern air lines. Special mention should be made of the important role played by radio in the newly established transpacific service by the Pan American Airways. Improvements have been made during the year in airway beacons and radio compasses; airport traffic control and blind landing systems are being tried out. In addition to the beacon and communication receiver, a small five-watt transmitter has been made available for the use of itinerant flyers in communicating with airports.

The use of radiotelephony with police cars is the most important application of radio with automobiles. There are two general types of this service, both of which have expanded materially during the past year. One consists of a one-way service from police headquarters to the cars and is usually conducted on a frequency in the range of 1500 to 2500 kilocycles. The other is a two-way service generally operating in the ultra-high-frequency range of 30,000 to 40,000 kilocycles.

[1] *Proc. I. R. E.*, March, 1936.

Another phase of radio with automobiles is the use of broadcast receivers in pleasure cars. Under-the-car antennas, made necessary by introduction of all metal automobile tops, inclusion of the radio control as a part of the instrument board, and the use of circuits for reducing ignition noise are the more important features of 1935 developments.

Photons and Electrons.[2] KARL K. DARROW. In a book entitled "Biological Effects of Radiation" edited by Professor B. M. Duggar of the University of Wisconsin, Dr. Darrow contributes the first chapter of 42 pages. In descriptive and brief manner the following topics are discussed: Waves and Corpuscles; Monochromatic light and measurement of wave-length; External photoelectric effect and measurement of photon energy; Units of wave-length, wave number, frequency, and photon energy; Regions of the spectrum; Absorption of light by atoms; Continua in absorption spectra, and ionization by light; Theory of absorption lines; Terms; Absorption in X-ray region; Emission of light; X-ray emission spectra; Production of X-rays; Production of light of the optical spectrum; Scattering of light without change of frequency; Scattering of light with change of frequency; Scattering of X-rays; Transmutation of electron-pairs and photons.

Neutralizing Transformer to Protect Power Station Communication.[3] E. E. GEORGE, R. K. HONAMAN, L. L. LOCKROW, E. L. SCHWARTZ. The use of commercial telephone circuits by power companies for a wide range of communication services including not only telephone but also telemetering, remote alarms, supervisory control and pilot wire control has focused attention on the problems of protection of this type of service. Where such circuits enter power stations which are subject to rise in ground potential at times of faults, the neutralizing transformer provides a means of securing adequate protection. Circuits operated into power stations through neutralizing transformers experience no adverse effects from potential rise up to 4,000 volts r.m.s. This result is produced by causing the transformer to introduce into affected communication circuits a counter voltage to neutralize the difference in ground potential. Transformers for indoor and outdoor use have been designed. The characteristics are such that they produce substantially no adverse reaction upon the transmission over the communication circuits they protect. Trials were made in the territory of the Tennessee Electric Power Company. In five locations

[2] Chapter in book, "Biological Effects of Radiation," Vol. I, McGraw-Hill Book Company, Inc., 1936.
[3] *Elec. Engg.*, May, 1936.

in which transformers have been installed, they have prevented interruption of the circuits not only for long periods but also for short periods lasting only for the duration of a surge.

On the Preparation of Iron and Steel Specimens for Microscopic Investigations.[4] FRANCIS F. LUCAS. A given lens system has certain potential resolving powers. This potential resolving power may or may not be fully realized in practice. Even a low power objective has remarkable resolving ability and the very high aperture objectives are capable of furnishing sharp brilliant images of details measuring about two hundred atom diameters.

The author describes in this paper methods and materials for the critical preparation of iron and steel specimens. A flotation apparatus which he has developed for the preparation of abrasives is described and a typical particle size analysis of a magnesium oxide abrasive prepared by this method is given.

Some Alloys of Copper and Iron (The Tensile, Electrical and Corrosion Properties).[5] EARLE E. SCHUMACHER and ALEXANDER G. SOUDEN. Bars of copper-iron alloy 0.75 and 1.0 in. in diameter and 20 in. in length were prepared with compositions ranging from 75 Cu–25 Fe to 37.5 Cu–62.5 Fe without segregation sufficient to detect by differences in electrical and mechanical properties. These alloys were hot worked satisfactorily to 0.25 in. rod. The copper-iron alloys in the range investigated consist of a mixture of solid solutions of the constituent elements, the phasial relationships of which depend on the thermal treatment.

A few of the observations made concerning these alloys are listed below:

1. The alloys in the range investigated are of the precipitation hardening type, but do not require a drastic quenching treatment to retain a supersaturated iron-rich phase. The optimum combination of tensile and electrical properties is obtained in the 50 copper-50 iron alloy by aging at 500° C. followed by hard drawing.
2. High tensile strengths associated with desirable electrical conductivities can be developed for certain of the compositions. An alloy of 50 Cu–50 Fe, for example, can be prepared in the No. 18 AWG with an ultimate strength of 180,000 to 190,000 lbs. per sq. in. and an electrical conductivity of approximately 30 per cent.

[4] *Trans. Amer. Soc. for Metals*, March, 1936.
[5] *Metals and Alloys*, April, 1936.

3. The alloy of 50 copper-50 iron can be satisfactorily tinned commercially.
4. Within the range of alloys studied, corrosion resistance decreases with increase of iron content. Various corrosion tests indicate that these alloys might prove corrosion resistant in inland rural districts, but that they are unsuitable for use in marine atmospheres and would probably be unsatisfactory in most industrial atmospheres, particularly in regions near the sea coast.

A Study of the Electromagnetic Field in the Vicinity of a Radiator.[6] F. R. Stansel. The complete equations for the electromagnetic field of an infinitesimal current element are given. The integration of these equations is considered for the case of a finite radiator having an empirical current distribution. Tables are included to facilitate computation and consideration is given to difference in phase of the current in various portions of the radiator.

An Analysis of Theater and Screen Illumination Data.[7] S. K. Wolf. During the past twenty years much information on theater and screen illumination has been accumulated. The significance and reliability of these data are discussed in the light of known physical factors influencing proper illumination. As a first approximation to a standard, it is suggested that the data indicate a value of 8 to 12-foot candles as representing satisfactory illumination. Variation of required illumination with screen size is analyzed, and a solution of the problem is suggested. The brightness of screen surroundings also is discussed. It is concluded that improvement in projection may be made by stricter application of existing information but that further investigations are desirable.

[6] *Proc. I. R. E.*, May, 1936.
[7] *Jour. S. M. P. E.*, May, 1936.

Contributors to this Issue

C. H. Amadon, B.S., Biltmore School of Forestry, 1908. Forester, Great Northern Paper Company, Maine, 1909–11; Forest Engineering, 1912–17; Western Electric Company, Inspection Department, 1918–24; Bell Telephone Laboratories, 1925–. Mr. Amadon has been engaged in development work in connection with timber engineering, timber utilization, and specification problems.

F. M. Bronson, Walton (N. Y.) Telephone Company, 1898–1902. New England Telephone and Telegraph Company, Engineering Department, 1902–04; Traffic Department, 1904–10. American Telephone and Telegraph Company, Department of Operation and Engineering, 1910–. Mr. Bronson has been engaged chiefly in work relating to methods and arrangements for handling toll traffic. He is now a Bell System representative in Washington, D. C.

R. H. Colley, A.B., Dartmouth College, 1909; A.M., Harvard University, 1912; Ph.D., George Washington University, 1918; Austin Teaching Fellow in Botany, Harvard University, 1910–12; Instructor in Botany, Dartmouth College, 1909–10 and 1912–16; Pathologist, Division of Forest Pathology, Bureau of Plant Industry, U. S. Department of Agriculture, 1916–28. Bell Telephone Laboratories, 1928–. Dr. Colley has been engaged in specification studies for poles and pole preservation research.

W. H. Doherty, S.B. in Electrical Communication Engineering, Harvard University, 1927; S.M. in Engineering, 1928. American Telephone and Telegraph Company, Long Lines Department, 1928. Research associate, radio section, U. S. Bureau of Standards, 1928–29. Bell Telephone Laboratories, 1929–. Mr. Doherty has been concerned with the development of high-power radio transmitters.

L. P. Ferris, B.A., University of Colorado, 1908; B.S. in Electrical Engineering, Massachusetts Institute of Technology, 1911. American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919–34, where he worked on inductive interference, crosstalk and measuring problems. Bell Telephone Laboratories, 1934–. As Inductive Coordination Engineer, Mr. Ferris has been concerned with questions as to the

effect of foreign voltages on telephone circuits, out of which arose the joint study of electric shock.

R. V. L. Hartley, A.B., Utah, 1909; B.A., Oxford, 1912; B.Sc., 1913; Instructor in Physics, Nevada, 1909–10. Engineering Department, Bell Telephone Laboratories, 1913–. Mr. Hartley has been associated with researches in the field of carrier current, telephone repeater, and telegraph systems.

L. W. Hussey, A.B., Dartmouth, 1923; M.A., Harvard, 1924; B.S. in E.E., Union College, 1930; Mathematics Department, Union College, 1924–29. Bell Telephone Laboratories, 1930–. Mr. Hussey has been engaged principally in work on the stability of regenerative systems and on modulation in non-linear resistances.

Robert N. Marshall, B.S. in Physics, Princeton University, 1930. Bell Telephone Laboratories, 1930–. Mr. Marshall has been engaged in acoustical studies pertaining to microphone design.

Edward C. Molina, Engineering Department of the American Telephone and Telegraph Company, 1901–19, as engineering assistant; transferred to the Circuits Design Department to work on machine switching systems, 1905; Department of Development and Research, 1919–34. Bell Telephone Laboratories, 1934–. As Switching Theory Engineer, Mr. Molina has made contributions to the theory of probability and its applications to telephone problems, such as the efficiency of various trunking arrangements and the significance of data derived from samples.

D. A. Quarles, A.B., Yale University, 1916; U. S. Army, 1917–19. Engineering Department, Western Electric Company, 1919–25; Bell Telephone Laboratories, 1925–. Mr. Quarles was earlier engaged in transmission studies of circuits and networks. More recently he was in charge of inspection engineering on apparatus products. As Outside Plant Development Director, he is now engaged in development work on outside plant products.

Frank F. Romanow, M.E., 1930, M.M.E., 1933, Polytechnic Institute of Brooklyn. Bell Telephone Laboratories, 1930–. Mr. Romanow has been engaged in the development of small condenser microphones and in research relating to acoustical instruments.

P. W. Spence, B.S., Columbia University, 1918; E.E., Columbia University, 1921. American Telephone and Telegraph Company, 1921–27; Bell Telephone Laboratories, 1927–. Since 1927, Mr. Spence has been largely occupied with studies on the effects of electric shock.

L. R. Wrathall, B.S., University of Utah, 1927; Graduate School, 1927–28. Bell Telephone Laboratories, 1929–. Mr. Wrathall has been engaged in the study of non-linear reactances.

Technical Developments Underlying the Toll Services of the Bell System

A Supplement to
The Bell System Technical Journal
July, 1936

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS
OF ELECTRICAL COMMUNICATION

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

PRINTED IN U. S. A.

Technical Developments Underlying the Toll Services of the Bell System

Table of Contents

Technical Developments Underlying the Toll Services of the Bell System*

Early Developments

General—Telephone Instruments

TELEPHONY involves the transmission of speech to a distance by electrical means. Speech itself, physically considered, consists of rapid longitudinal variations in air pressure, or acoustic waves as they are called, traveling out from the mouth of the speaker or to the ear of the listener. Each sound has its characteristic wave form or group of wave forms and as a result these acoustic waves are of complicated and rapidly changing wave form as is illustrated by the oscillograms on Fig. 1 showing the structure of the electrical current

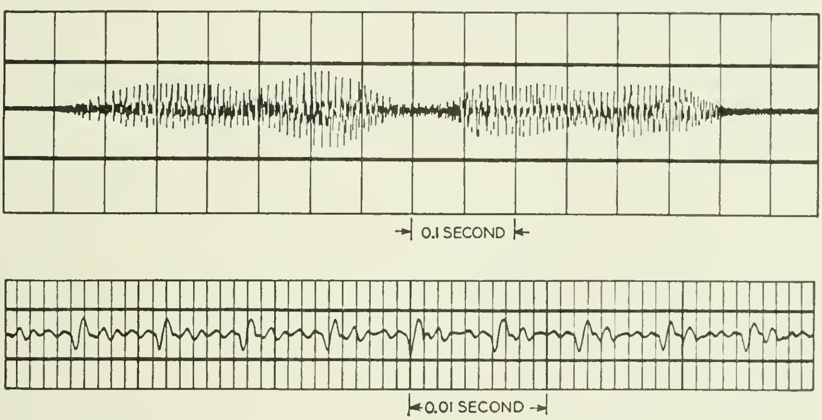


Fig. 1—Oscillograms showing electrical current in a telephone circuit resulting from spoken word "Harvard" and vowel in "Har."

in a telephone circuit resulting from the spoken word "Harvard," and, in more detail, the vowel in "Har." The telephonic transmission of speech requires, therefore, three fundamental elements:

(*a*) An instrument which, when acted upon by the acoustic waves of the speaker's voice, produces in an electrical circuit oscillations or waves suitable to represent the voice of the speaker. This is the telephone transmitter.

*At the request of the Federal Communications Commission, this pamphlet was recently prepared to give the Commission a brief account of some of the principal technical developments which have led to telephone toll service as given by the Bell System. As it brings together in concise form a summary of a large amount of information of interest to communications people generally, it is being issued with a few minor editorial revisions as a supplement to the Technical Journal.

H. S. Osborne

(*b*) A circuit from the point of transmission to the point of reception suitable to transmit these complicated electrical oscillations in the proper magnitude without undue distortion of form and without interference from electrical currents from other sources. This is the telephone circuit.

(*c*) An instrument which, receiving the electrical oscillations transmitted over the line from the telephone transmitter, reproduces acoustic waves of proper loudness and quality to correspond with those produced by the speaker's voice, by means of which waves the speech is transmitted to the listener. This is the telephone receiver.

The telephone invented by Bell in 1875 corresponded in principle to the telephone receiver of today and could be used alternately as a telephone transmitter and a telephone receiver. It consists of a diaphragm of magnetic material associated with a magnet and coils of wire. When this instrument is placed before the speaker's mouth, the variations in acoustic pressure cause the diaphragm to vibrate. The instrument is so designed that this vibration in the presence of a magnet produces electrical oscillations in the coils of wire. When these oscillations are transmitted over the telephone circuit to the coils of wire in a similar instrument, they cause variations in the strength of the magnet which, in turn, cause vibrations of the diaphragm of the receiving telephone. The vibrations of the diaphragm produce acoustic waves which reproduce the speech of the talker at the distant end of the circuit.

This instrument is very inefficient as a telephone transmitter and from earliest days efforts were directed toward the development of transmitters working on a different principle. Bell himself suggested the principle most generally used. This principle is that the vibration of the transmitter diaphragm shall vary the resistance of a local electrical circuit through which current is caused to flow by a battery. The variation in resistance can cause variations in the flow of current sufficient to induce relatively powerful electrical oscillations in the telephone circuit—in fact, the oscillations so induced may have a power several hundred times as great as that of the acoustic waves produced by the speaker. The telephone transmitters acting on this principle are, therefore, powerful amplifiers.

In his early work, Bell devised a transmitter working on this principle consisting of a small platinum wire attached to a diaphragm of goldbeaters skin and dipped very slightly into acidulated water held in a conducting cup. It was with this instrument that the first

complete telephone sentence, "Mr. Watson, come here, I want you," was successfully transmitted on March 10, 1876.

Almost from the first, the efforts of inventors to develop successful telephone transmitters made use of this principle, and while many variable resistance elements were tried out with some degree of success, transmitters employing granular carbon, the resistance of which varies with pressure, were the most satisfactory. Such a transmitter devised in 1878 by Hunnings of England using powdered "engine coke" was extensively used commercially. Better performance was provided by the design in 1890 by White of the Bell System of the so-called "solid-back" transmitter. This principle and the use of carbon transmitter elements have survived through numerous improvements in transmitter design and are applied to millions of telephone instruments in use today.

The telephone station of today includes, generally speaking, a transmitter of the variable resistance type, a receiver based on the principle of Bell's original discovery, both of these instruments being modern in design, includes induction coils, condensers, etc., necessary for electrically associating the transmitter and the receiver with each other and with the telephone line and, in addition, includes such items as a bell, a switchhook, etc. which are necessary for signaling and control of the telephone circuit.

Telephone Switching Systems

Very early in the practical use of the telephone, it became evident that the full usefulness of this method of communication required the development of means by which any subscriber could quickly obtain connection between his telephone and any other telephone rather than being limited in his conversations to one other subscriber or a small group of other subscribers connected together on the same telephone circuit. The difficulties which would be encountered with a telephone plant consisting of large numbers of stations connected to one circuit are obvious, the outstanding disadvantage being that only two subscribers could carry on a conversation over the circuit at one time. These difficulties led to the development of telephone switchboards at which connection could be made between lines to any two subscribers in a given town or city.

As technical developments made toll service between different cities possible, means were needed for the rapid connection of any two subscribers in different cities. It would obviously be impracticable to connect together at the same switchboard subscribers in distant cities, and switching systems were adopted so that toll connections between any two subscribers in different cities could be established

over telephone lines terminating in toll switchboards located in those cities, and trunks between the subscriber switchboards and the toll boards.

As the number of subscribers and the extent of telephone service increased, it became impractical and uneconomical to connect all telephone subscribers in the larger cities to the same switchboard; impracticable because the size of such a switchboard would be so great as to make the interconnection of two lines an unwieldy and slow procedure; uneconomical because of the relatively large amount of telephone line which would be required to connect the more distant subscribers with the central office. For these reasons means of interconnecting switchboards within a city were devised whereby a station terminated on one switchboard can be connected to a station terminated on another switchboard in the same city over a telephone line or "interoffice trunk" terminating on each switchboard. The design and layout of the subscriber and switchboard plant require careful consideration in determining the maximum economy which can be realized with the proper balance between subscriber lines and interoffice trunks.

Telephone Circuits and Cables

At the beginnings of telephone service it was found that the iron wire then used for telegraph circuits was, in many cases, not satisfactory for telephony because of the losses of energy taking place in the wires and the rapid diminution in the loudness of the transmitted speech with the distance over which it was transmitted. At first no wire was available having better electrical characteristics and at the same time sufficient mechanical strength to withstand the strains it was subjected to when strung on a pole line. Thomas B. Doolittle of the Bell System, who was familiar with certain physical properties of copper, conceived that if copper were drawn cold through a series of dies, he might obtain a wire of much greater physical strength than the soft annealed copper wire then used in a small way in the making of electrical apparatus. In November, 1877 he arranged with a manufacturer to try the process and it was so successful that in 1878 a quantity of hard-drawn copper wire was placed in service in the Bridgeport, Connecticut exchange. The success of this and subsequent installations showed that a wire which was electrically efficient and mechanically strong had been obtained by means of which telephone service could successfully be given over considerable distances.

The numbers of wires required to serve telephone subscribers in large cities led at an early date to the development of means for putting

the wires underground. The early experiments starting in 1878 took advantage of the known advantageous properties of lead water pipes. Insulated copper wires were drawn into lead pipes of this character. By 1886, practical means had been developed whereby lead heated to the point of plasticity could be extruded over a compacted group of insulated conductors, thus forming the pipe tightly about the conductors, and this general principle has been followed in telephone cables to the present day. By 1890 there was a general development of insulated telephone cables in the congested parts of the larger cities. In 1891 Bell System engineers introduced the use of paper for insulating cables, and this practice is still followed in cable manufacture.

Early Toll Service

The success of the early installations of hard-drawn copper wire for short lines indicated that a type of conductor had been developed by which it might be possible to extend telephone service over considerable distances. This led to a very important experiment, the construction in 1883 of an experimental toll line between New York and Boston carrying two wires.

Prior to the construction of the New York–Boston line, telephone lines, following telegraph practice, were generally of one wire grounded at both ends, the so-called "ground return circuit." However, based upon experiments with iron wires over shorter distances, particularly between Boston and Providence, J. J. Carty of the Bell System had determined that the ground return circuit was so noisy, due to interference from telegraph lines and other causes, that such circuits could not be used over long distances, and had also discovered that by using two wires connected as a metallic circuit, the interference was very greatly reduced. Carty's metallic circuit was used with success in the New York–Boston line and was adopted for all the following construction of long toll circuits. The metallic circuit principle thus developed first for toll lines extended back into local lines so that in highly developed areas all telephone circuits are now constructed on the metallic circuit principle.

This experiment successfully demonstrated the practicability of "long distance" transmission and led to the determination to extend long distance service as widely and as rapidly as the state of the art permitted and to the incorporation in 1885 of the American Telephone and Telegraph Company for this purpose. The first telephone line constructed by this new company was the New York–Philadelphia line, using hard-drawn copper wire on a metallic circuit basis. It was found that with two or more metallic circuits on a pole line, speech

currents flowing in one circuit will cause similar, weaker currents to flow in the other circuits. This is called induction or crosstalk and experience with the New York-Philadelphia line showed that so much induction between telephone circuits was obtained that intelligible crosstalk resulted; in other words, one could overhear on one circuit what was said on the others. To overcome this, systems were developed whereby, by suitably interchanging positions of the wires of a circuit, the inductive effects in that circuit from an adjacent circuit would tend to neutralize. This is illustrated in Fig. 2 for a simple case of two circuits. In this figure the circuit sections "*a*" and "*b*" are equal in length, and voltages are being induced into circuit No. 2 from circuit No. 1. The arrows show the directions in which the induced current would tend to flow in circuit No. 2, the wires of which are interchanged in position between the two sections. It will be noted that the induced voltage in section "*b*" is equal in magnitude to that in section "*a*" and, by the interchanging of wires at the

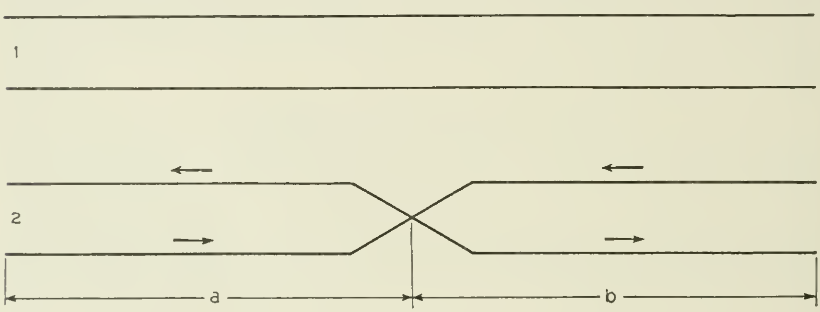
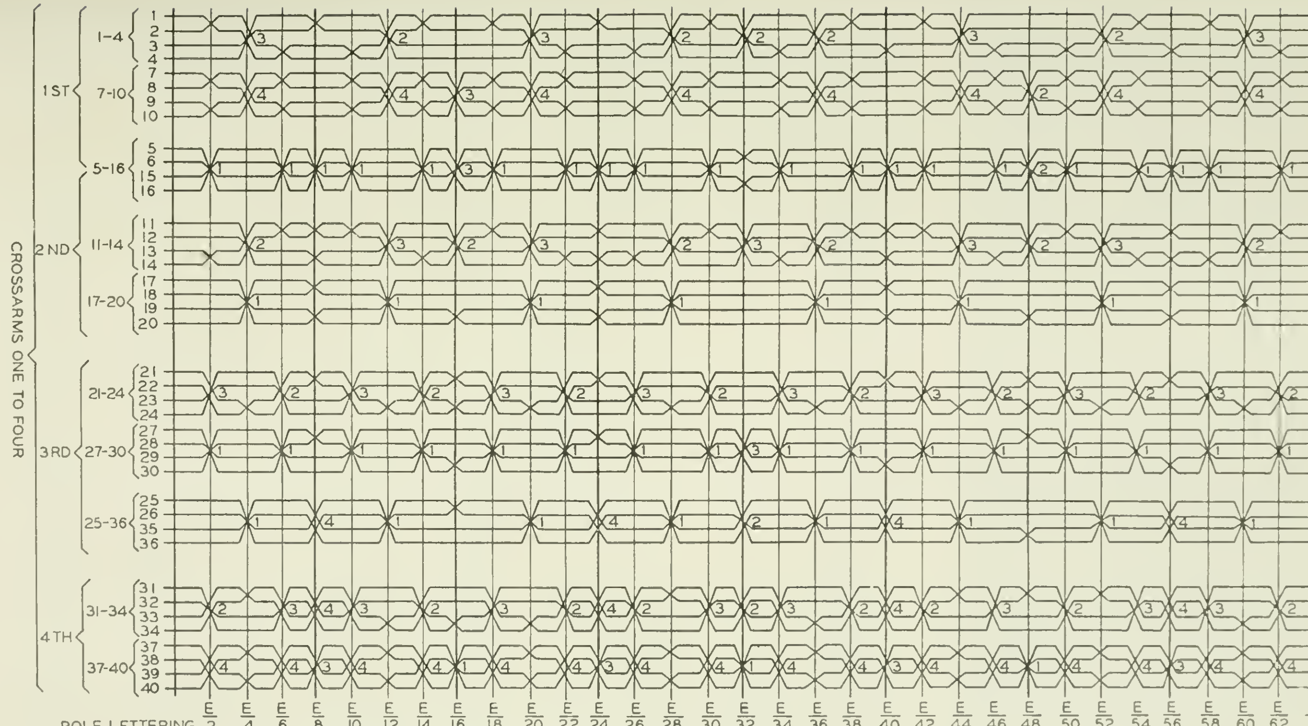


Fig. 2—Side or non-phantomed circuit transposition.

junction, is made to oppose that of section "*a*." The effect of this interchange or transposing of wires is such as to neutralize the induction in sections "*a*" and "*b*" appearing at the circuit terminals.

With a large number of circuits, the induction between each two circuits must be neutralized in each short section of line, and to accomplish this, more complicated arrangements, known as transposition systems, were developed. The first system of this sort was worked out by J. A. Barrett of the Bell System in 1886. The development of transposition systems has continued constantly since that time, the problem changing with the increase in the number of circuits on a line, developments in the transmission of electrical power on lines which sometimes are constructed near the telephone lines, the introduction of phantom telephone circuits, and of repeaters and carrier telephone circuits into the plant. Figure 3 shows a typical transposition system in use in the Bell System today.



NOTES: – LINE TRANSPOSED IN DIRECTION OF ARROW ⟶
THE FIGURE AT EACH PHANTOM TRANSPOSITION INDICATES THE TYPE. SPACE AS REGULARLY AS POSSIBLE

Fig. 3—Typical transposition system in use in Bell System today.

With the more extensive application of the cable developments mentioned above to local circuits, it was natural that they should be extended to the longer circuits used for toll business. In 1898 a 30-pair 16-gauge cable insulated with paper was extended eight miles from Boston toward Lynn, Massachusetts. Shortly after this a 30-pair 14-gauge cable was placed between Boston and Wakefield, Massachusetts, a distance of about 12 miles. It was found, however, that with the increasing distance in cable, the loss in transmission rapidly increased since cable circuits, because of the small size of the wires and the large electrical capacitance, had inherently poorer electrical characteristics for the transmission of telephone currents than the larger open wires strung on poles.

The Phantom Circuit

The phantom circuit has grown out of a conception of Jacob in 1883 which is illustrated in principle in Fig. 4. He conceived that by

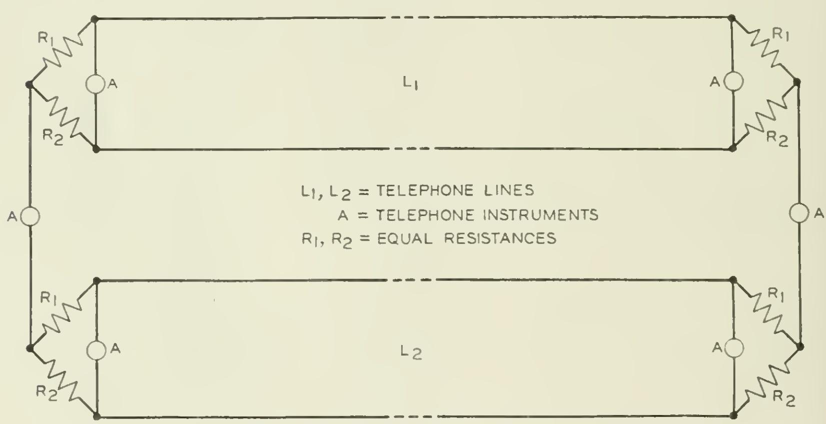


Fig. 4—Phantom circuit—conception of F. Jacob.

bridging resistances across each end of two parallel telephone circuits, a third circuit could be created as indicated by connecting telephones at each end between the midpoints of these resistances. These three telephone circuits could, therefore, use four wires without mutual interference. While this scheme was not practicable, it led to a proposal by Mr. Carty in 1886 to substitute balanced transformers (called repeating coils) in place of the resistances as indicated in Fig. 5.

In order to successfully apply this idea it was necessary to develop repeating coils that were carefully balanced, that is, which had the two halves of their windings very exactly equal in electrical characteristics, so that the current from the phantom circuit would divide equally between these two halves of the windings and not influence the other circuits (called the "side" circuits). Also, an improved technique of line construction was necessary in order to avoid high resistance joints and other irregularities in construction which would result in overhearing between the phantom and the side circuits. Furthermore, in order to avoid overhearing between different phantom circuits on the same pole line, it was necessary to interchange not merely the two wires of each pair but also all four wires of the phantom

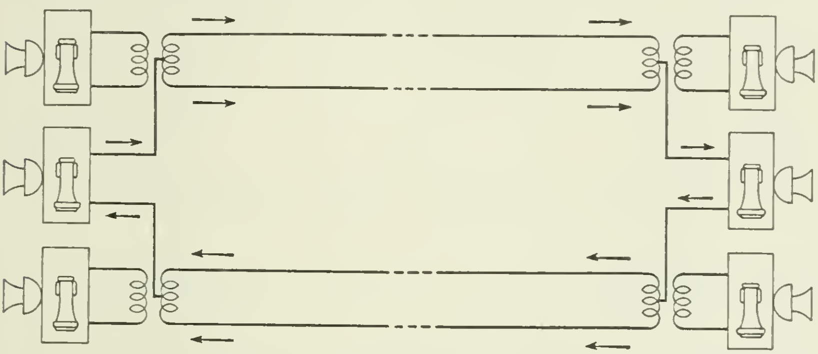


Fig. 5—Diagram of phantom circuit using balanced transformers (called repeating coils).

group as is indicated in the transposition system shown in Fig. 4. This greatly complicated the design of transposition systems. An important feature of the problem is the high degree of balance required since the transfer from the phantom circuit to the side circuit of more than one-millionth of the electrical energy carrying the telephone currents in the phantom circuit or vice versa might be sufficient to make overhearing possible. This high degree of balance was achieved by years of painstaking work and resulted in the first successful phantom circuits in the year 1903.

Today 12,400,000 miles of wire in the Bell System are installed in such a way as to be suitable for phantom operation. Without phantoming, 6,200,000 additional miles of wire would be required for the same circuit mileage.

Superposed Telegraph on Telephone Circuits

From the very beginnings of long distance telephony, the telephone wires were used also for private line telegraph service. At first,

means had not been developed for using the wires simultaneously for both telephone and telegraph, and the two services were offered alternatively to the private line customers. Beginning in 1887, however, successful experiments were conducted in using telephone wires simultaneously for telephone and telegraph services by the method of superposition which is shown in Fig. 6.

The first method, called "simplexing," is an adaptation of the phantom principle for the use of telegraph on telephone circuits, the grounded telegraph circuit being introduced at the midpoint of repeating coils at the two ends of the telephone circuits, the currents dividing equally in opposite directions so that there is no interference between the telephone and telegraph circuits. The other method of simultaneous operation of telephone and telegraph circuits, however,

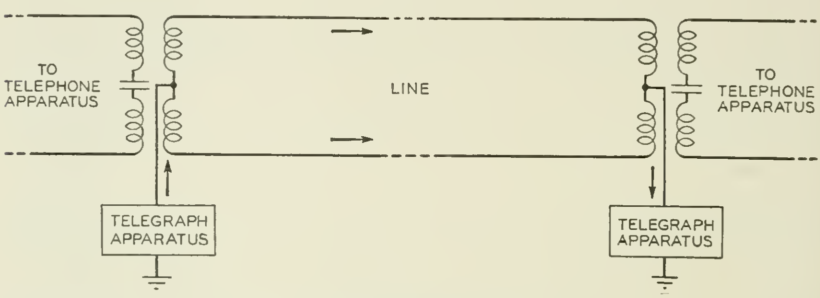


Fig. 6—Schematic of telegraph circuit superposed on a telephone circuit using the "simplex" method.

depends upon a new principle and one which has come to be of the greatest importance in the subsequent development of telephony. This principle is the selection and separation of electric currents into different channels depending upon differences in their frequency of alternation.

While the form of the electrical oscillations which transmit speech over a telephone circuit is extremely complicated, as indicated in Fig. 1, such oscillations can, by processes of analysis, be considered as made up of a large number of simple alternating currents of different frequencies. A simple current of this type, which is sometimes spoken of as a sine wave because of the mathematical law which expresses the variation in the flow with time, is shown in Fig. 7. Such a current by gradual variations at regular intervals reverses its direction of flow. Each double reversal is called a "cycle" and the number of such double reversals in a second is called the frequency of cycles per second.

An analysis of telephone currents shows that, in order to transmit satisfactory speech, it is necessary for all of the telephone apparatus and circuits involved to transmit with nearly uniform efficiency simple alternating currents over a considerable range of frequencies. For new designs of telephone circuits, the minimum range so transmitted is between about 250 and 2,750 cycles. The voice contains components of lower frequencies and also of high frequencies but it is not necessary to transmit them because their contribution to the clearness of the speech is relatively unimportant.

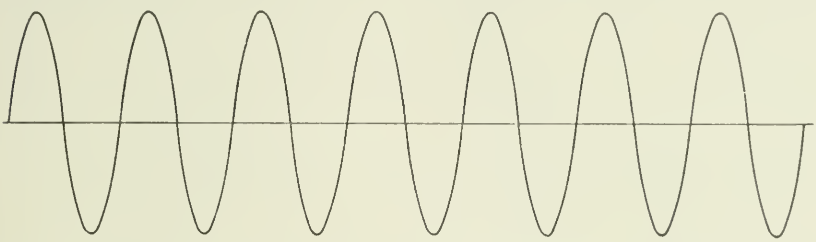


Fig. 7—Graph of a simple alternating current or "sine wave."

A similar analysis of the currents used for telegraphic transmission shows that they may be considered also as composed of simple alternating currents covering a band of frequencies—with equipment generally used in the Bell System, this band extends from zero up to roughly about 100 cycles. Components of the frequencies above about 100 cycles can be excluded from the telegraph circuit without reacting upon its efficiency of transmission with the equipment and the speeds of signaling commonly employed in private line telegraph circuits. This difference in the range of frequencies required for the transmission of telephone currents and for the transmission of telegraph currents makes possible the application of the principle of separation of electrical currents into different channels depending on the difference in their frequency of alternation mentioned above. Apparatus placed at the terminals of the circuits, which is called "composite sets," is so designed that telegraph currents and the telephone currents can be transmitted into the same wires and at the receiving end are separated into the telephone and telegraph channels, respectively, without interference. The form of this apparatus is indicated diagrammatically in Fig. 8.

The principles of simplexing and compositing have been applied extensively to the long distance circuits of the Bell System, there being now in service approximately 760,000 miles of telegraph circuit oper-

ating on these principles using wires simultaneously with their use for telephone transmission without mutual interference.

Development of the Mathematical Theory of Transmission—Loading

As telephone lines came to be extended over greater distances, it was evident that, even with the best copper telephone circuits, the loudness of speech transmitted over the circuit rapidly became less with distance, and also, particularly when the circuits were in cable, the clearness of the speech was impaired at the greater distances. At first these effects were not clearly understood, there being no adequate quantitative analysis of the effects on telephone transmission of the various electrical characteristics of the telephone circuits. Throughout all the early development period, the continued study of the

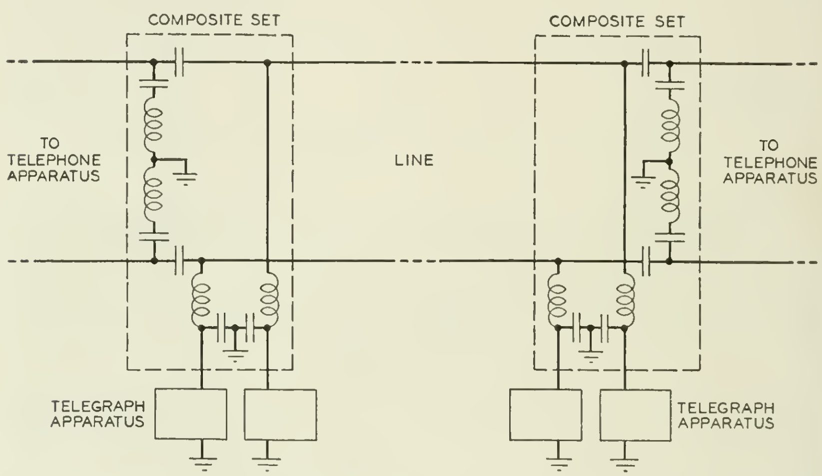


Fig. 8—Schematic of a telegraph circuit superposed on a telephone circuit using "composite sets."

mathematical theory of the transmission of currents over wires led to an increasing insight into these problems and into the conditions necessary for transmitting telephone currents over long distances efficiently and without undue distortion. The foundation was laid in the masterful, if sometimes enigmatic, papers of Oliver Heaviside published over a long period of years beginning with 1882. One result of Heaviside's work was an appreciation on his part of the unexpected fact that an improvement in transmission efficiency of telephone circuits would be brought about by an increased inductance of the telephone circuits, and he suggested in his papers that means

might be found to increase the inductance artificially. This suggestion was taken up by two investigators in America, Professor M. I. Pupin of Columbia and Dr. G. A. Campbell of the Bell System who, working independently, proved by further mathematical development that this could be done practically and showed how to do it. These mathematical studies showed that, while the addition of inductance in large quantities at one or several points in the circuit destroyed its capability for transmitting telephone currents, the insertion of inductance in smaller quantities at regular and frequent intervals by means of highly efficient inductance coils would greatly improve the transmission efficiency.

Interurban Toll Cables

The development of practical means of applying the loading principle had been stimulated by the need for some practical means of improving the efficiency of toll cables. This principle led promptly to the extension of interurban toll cables, important items being the installation in 1906 of cables between New York and Philadelphia, a distance of about 90 miles, and between New York and New Haven, a distance of about 80 miles, and, in 1908, of a cable between Chicago and Milwaukee having a length of about 90 miles. At about this time experimental work was being actively conducted by the Bell System in the effort to develop a type of construction for toll cables which would permit the use of phantom circuits in the cables as well as in open wire. This required a new technique in cable construction, involving new principles and many refinements in detail to eliminate the interference which would exist between phantom circuits and the side circuits from which the phantoms are derived and also between phantom circuits in the same cable. The processes worked out included not only manufacturing methods but new types of electrical tests and new splicing procedures applied in the course of installation by means of which the unbalances in successive lengths of cable are made to largely neutralize each other. As a result of this work, a successful phantom cable was installed between Boston and Neponset, Massachusetts in 1910, a distance of six miles.

This work led to the inauguration in 1911 of a very important interurban cable project. At the time of the inauguration of President Taft on March 4, 1909, a sleet storm of unprecedented severity had broken down all the wire lines entering Washington and isolated the Capitol from the rest of the country. The Bell System management determined that, as soon as technical advances made it possible, means would be adopted for insuring against any future similar interruption

of the communications between the United States Capitol and the rest of the nation. Upon the success of the experiments described above, it was decided to complete an underground cable route connecting Washington with Baltimore, Philadelphia, and New York, using large gauge conductors, the phantoming principle which had just been successfully demonstrated, and new systems of loading designed specifically for the new cable. The project was completed in 1912 and in 1913 this high grade cable route was extended to Boston, through New Haven, Hartford and Providence.

Possibilities and Problems Associated with the Use of the Telephone Repeater

The developments discussed above had done a great deal to extend the range of telephone service making possible good commercial service between the Atlantic Seaboard and Chicago and a service of a kind as far west as Denver and providing a storm-proof cable route connecting Washington and Boston and the intermediate cities of the Atlantic Seaboard. By 1912, however, it was apparent that in addition to pushing to the utmost the advantages to be gained from the technique already developed, it would be necessary, if universal service for the entire country were to be realized, to find satisfactory means for amplifying the attenuated telephone currents on a long telephone circuit so that after transmission over one section of line they could be restored, in amplitude, transmitted into a second section, and when again attenuated restored a second time and transmitted into a third section of the line and so on, without undue distortion or change in the structure of the voice currents. The device to accomplish this is called a telephone repeater. The conclusion that improved repeaters were required was reached after a careful analysis of all of the possible means of achieving further extensions in the range of long distance transmission and as a result the energies of the research forces of the Bell System were to a greater extent than before directed to the development of improved telephone repeaters and of circuits and methods of line construction which would make possible their general use.

One of the chief problems which confronted the engineers undertaking the intensive telephone repeater development work beginning in 1912, was the development of an amplifying element for a repeater which could be used generally for telephone purposes. The telephone repeater was not new in the art at that time, since a repeater giving beneficial results had been invented by H. E. Shreeve of the Bell System and first used successfully on a circuit between Amesbury, Massachusetts and Boston in 1904. The Shreeve repeater took ad-

vantage of the amplifying characteristics of a variable resistance telephone transmitter and combined in one instrument, in refined form, the fundamental elements of a telephone receiver and a transmitter. The attenuated telephone currents entering the receiver side of the device caused the vibration of a diaphragm which, in turn, actuated a variable resistance element which transmitted amplified currents to the next section of telephone line. Figure 9 shows a cross-section of the amplifying element of this repeater, commonly known as the mechanical repeater. Repeaters of this type were used in commercial service for a number of years but since development

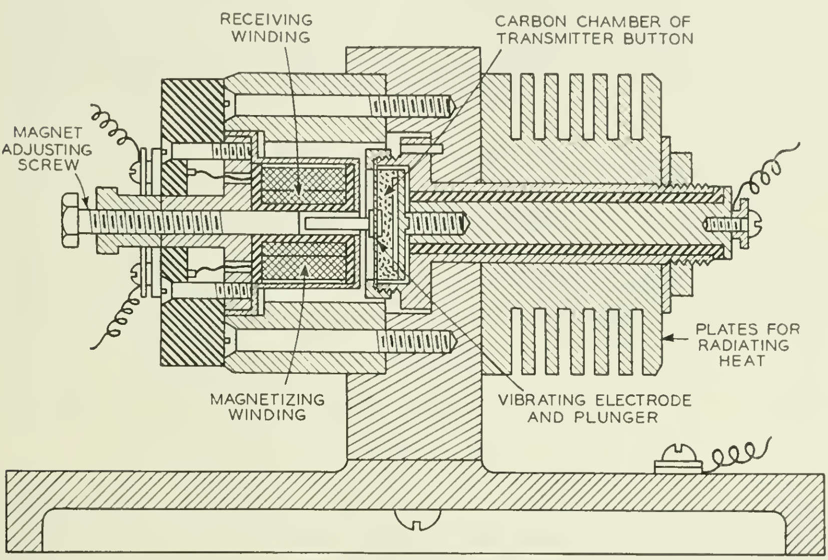


Fig. 9—Cross-sectional drawing of a mechanical type of telephone repeater.

work up to the time of the application of vacuum tubes to telephone repeaters had not overcome the fundamental difficulties of distortion, gain limitations and instability of the mechanical repeater, its use was gradually discontinued after the introduction into the telephone plant of the vacuum tube repeater. Other arrangements such as variation of the field current of a generator to produce corresponding variations in the armature voltage and the electromagnetic control of gaseous devices were tried out but were never successfully applied in any important degree to telephone circuits.

In 1906, Lee DeForest demonstrated before the American Institute of Electrical Engineers that the discharge between a hot cathode and a plate of a thermionic tube can be controlled by a third electrode.

Immediate use of this discovery was made in improving radio telegraph receivers. The tubes and circuits thus employed were not of types that could satisfactorily be used in telephone work which required a high stability of the amplifying device and freedom from distortion of speech currents. However, an intensive study of the possibilities of this device showed that the use of such tubes based on DeForest's discovery was by far the best method of amplifying telephone currents yet developed.

Development work on vacuum tubes carried on by the Bell System has included the perfection of the tubes by the use of high vacuum, scientific proportions and new types of filaments to secure improved efficiency. The performance of vacuum tubes used in the Bell System has been improved extensively through continued development work. For example, during the last twelve years the average life of the tubes used in the Bell System has been extended by a factor of 10, and, at the same time, their power consumption has been reduced appreciably.

Vacuum tubes were first applied to telephone repeaters experimentally, and to a small degree commercially as early as 1913. One of the first important uses of vacuum tube repeaters, however, was in 1915 in connection with the first transcontinental telephone service between New York and San Francisco, a distance of approximately 3400 miles. The circuit consisted of No. 8 B.W.G. open-wire copper conductors loaded at eight-mile intervals and having vacuum tube telephone repeaters located at Pittsburgh, Omaha, and Salt Lake City.

Years of experience with early forms of telephone repeaters had shown that the successful use of repeaters depended not only upon the development of a suitable amplifier but also upon the design of suitable circuit arrangements for associating the repeater with the telephone line and on improved methods of line construction. An important consideration is the fact that a telephone circuit must operate in both directions, that is, it must permit talking to be carried on from either end of the circuit. A single telephone repeater element, however, is inherently a one-way device, receiving attenuated currents at one pair of terminals and transmitting amplified currents from the other pair of terminals. The association of such one-way elements with a two-way telephone circuit is not a simple matter because if any considerable proportion of the amplified output current of the repeater reaches the input terminals, it is again amplified and results under ordinary conditions in turning the repeater into a generator of alternating currents (an oscillator) and destroying its usefulness as a repeater.

The type of circuit arrangement most commonly used to associate two repeater elements with two sections of telephone line in such a way that the operation will be satisfactory is shown schematically in Fig. 10. This is known as a 22-type repeater. Attenuated telephone current transmitted from a distant point over the west section of line passing through the transformer *A* of a special design (sometimes called a hybrid coil) enters the input of amplifier element *B* and is amplified. From the output of amplifier element *B* it passes through a second transformer *C* associated with the east section of line and a balancing network *E*. An essential function of the repeater circuit lies in the design of the transformer in such a way that the two halves (between the midpoints of which the input to the amplifier element *D* is connected) are equal and in the design of the east line and of the balancing network *E* in such a way that they offer the same impedance to the

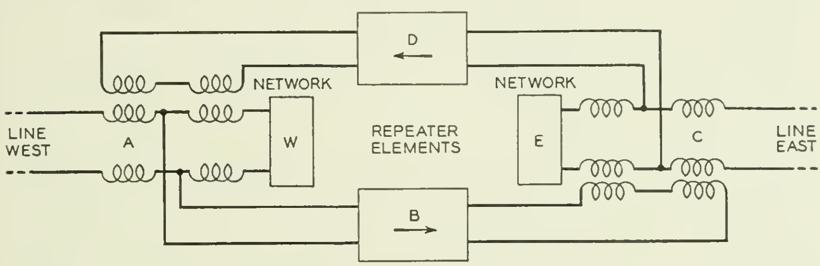


Fig. 10—Schematic 22-type telephone repeater circuit.

flow of current. Coil *C* is so designed that if this condition is exactly met, none of the ouput current from amplifier element *B* is transmitted across to the input element of amplifier *D*. However, currents transmitted in from the east section of line pass through Coil *C* to amplifier *D*, are amplified and retransmitted to Coil *A* where the condition of balance also must be applied between the west line and the balancing network *W*.

The complete repeater circuit includes many other things beside the bare essential elements shown in Fig. 10. Important among these are electrical filters to control the band of frequencies, potentiometers to control the amplification, transformers to efficiently interconnect different parts of the circuit, equalizers (see chapter on Associated Technical Developments) and arrangements for the supply of electric power to the vacuum tubes.

Obtaining a condition of balance requires that each section of line offer the same impedance to the flow of current as the balancing net-

work associated with that section of line in the repeater circuit. A difficulty to be met arises from the fact that telephone currents are very complicated in wave form as previously indicated and involve components varying all the way from 200 or 300 to 2700 cycles per second or more. In order to provide for suitable operation, the condition of balance must be met within a close approximation for currents of all of the frequencies within this range. The reason why this requirement reacts on the construction of telephone lines is very simply illustrated by the curves of Figs. 11 and 12. Figure 11 shows the impedance of a long telephone circuit for all frequencies within that range when the circuit is of very uniform construction throughout. Figure 12 shows the corresponding impedance curve obtained if there are some irregularities in construction in the line. It is possible to design balancing networks which have the same characteristics as those indicated in Fig. 11 for the uniform line but it is not practicable without too great expense to design such networks having the same characteristics as the irregular line shown in Fig. 12. This is true since the characteristics of no two irregular lines are the same, the characteristics varying widely depending upon the nature and the location of the irregularities.

Means for the general use of repeaters on telephone circuits therefore involved the development by the Bell System of line balancing networks and the development of long telephone lines with uniform impedance characteristics over the range of frequencies used in telephony. In some cases this could be done by making the lines uniform in construction. For example, loading coils had to be designed so that they had, very accurately, equal amounts of inductance and had to be spaced at exact equal intervals along the line. Furthermore, it was found that the types of loading coil in previous use were affected by lightning and other causes so that the amounts of inductance changed enough to interfere with repeater operation. It was, therefore, necessary to develop new types of coils of very stable materials which would avoid this change in inductance.

In some cases, it is not possible to construct the line in a uniform way throughout. For example, it is often necessary for open-wire lines to be brought into towns and cities through sections of cable. For such cases, for each type of open-wire construction, a type of cable construction was worked out having such characteristics that it could be connected to the open wire without spoiling the impedance characteristics of the circuit. This involved the development of new loading systems for use on cables of this sort. In the case of circuits entirely in cable, improved uniformity in the manufacture of the cable

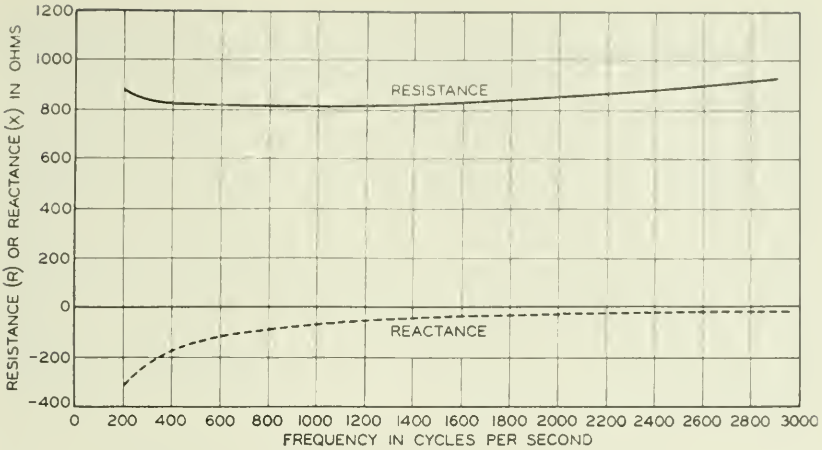


Fig. 11—Impedance-frequency characteristics of a smoothly constructed telephone line. Impedance $(Z) = \sqrt{R^2 + X^2}$.

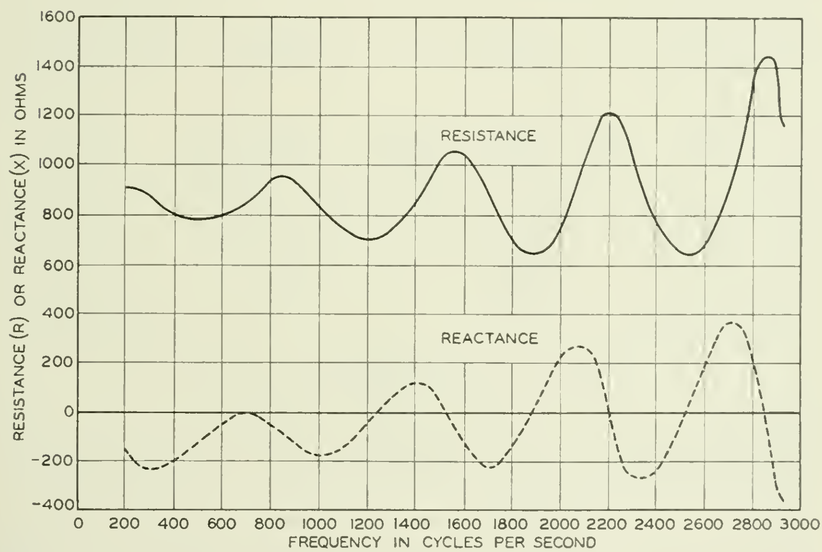


Fig. 12—Impedance-frequency characteristics of a telephone line having an impedance irregularity.

itself was required as well as improved loading coils and more exact rules for their location.

Generally speaking, repeaters are used in connection with the phantom circuit arrangements previously described, by means of which three independent telephone circuits are derived from four

wires. The application of repeaters to such a group of four wires in cable (spoken of as a phantomed group or a quad) is shown schematically in Fig. 13. In this figure, the boxes denoted "Telephone Repeater" represent the complete repeater circuit shown in Fig. 10. It is necessary to separate the telephone currents of the phantom circuit from those of the two side circuits by applying to the phantom group highly balanced repeating coils, just as is done at the terminals of the circuit, and providing separate repeaters for each of the two side circuits and the phantom as is illustrated in the figure. The figure also shows a typical telegraph circuit arrangement—a metallic telegraph circuit on each pair, separated from the telephone channel by composite sets as previously described, and passing through telegraph repeaters at the telephone repeater point.

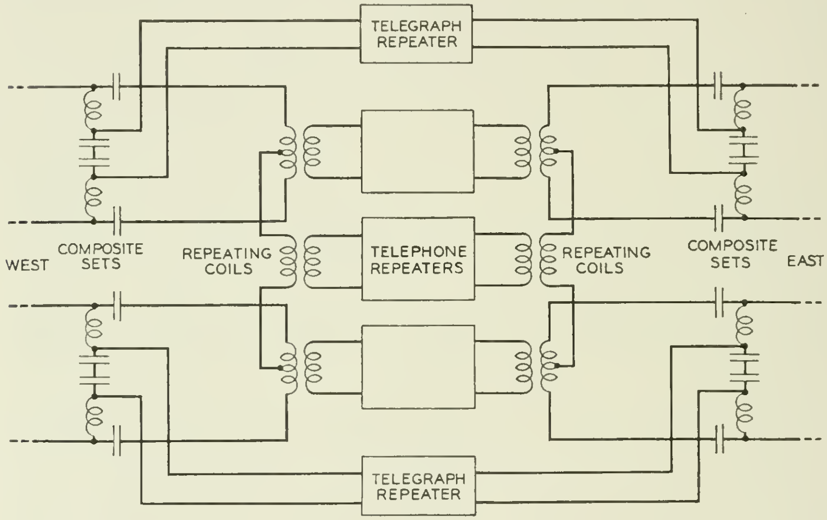


Fig. 13—Showing schematically the association of the four wires of a phantom group and composite sets, repeating coils, telephone repeaters and telegraph repeaters.

Phantom operation through cables, without crosstalk between the phantom and its side circuits and between the circuits in different quads of the cable, has involved a long train of developments in decreasing the tolerances of manufacture and increasing the uniformity in the characteristics of the cable. The cables must meet a double requirement, namely, freedom from crosstalk, which is made more difficult with the application of repeaters, and uniformity of characteristics to provide for suitable repeater operation. This double requirement has been met by niceties in design, construction, and

installation, including a series of delicate electrical tests on the cable and portions of the cable during the process of installation.

With these developments, it became possible to use repeaters to provide a large improvement in the transmission efficiency of telephone circuits both in open wire and in cable, a number of repeaters where necessary being used at different points on the same circuit. The repeater element itself could be made so free from distortion that a very large number of them could be used in succession on the same circuits. The limitation in the use of repeaters, however, was determined largely by the degree of balance practicably obtainable through the more uniform construction of the telephone lines. While it was practicable to have such a degree of balance that a single repeater would amplify the telephone current often six or sevenfold, it was not generally practicable even with the use of a number of repeaters to extend the length of circuit using a given size of conductor over eight to tenfold. For transmission over very long distances or for the use of very small conductors in long toll cables, another principle was developed which was applied to toll cables and to the use of carrier systems on open wire, and will be discussed in connection with those subjects.

Toll Cable System

The success achieved in the general application of repeaters to telephone toll circuits opened the way for a great extension in the use of toll cables. Toll cables have the obvious advantage of providing a high degree of security and continuity for telephone circuits due to the fact that they are nearly immune from the effects of bad weather, particularly of sleet and of high winds which sometimes seriously interrupt open-wire telephone service by placing on the conductors and on their supporting structures loads greater than those for which they can economically be designed. Also, cables form the practical means for providing the very large numbers of circuits required to take care of the demand on very heavy routes by making it possible to crowd into one route a much greater number of circuits than could be provided by the ordinary open-wire technique.

Before the general use of repeaters became practicable, toll cables had the inherent disadvantage that, with the small conductors necessary to place a large number of circuits in one cable (until recently maximum outside diameter $2\frac{5}{8}$ inches), the cable circuits, even when equipped with loading coils, had a very high attenuation loss per mile compared with the open-wire circuits. Even when very large conductors were used in the cable at the sacrifice of the number of circuits

in order to provide circuits of high efficiency, as was done in the first cable between Washington, New York, and Boston, the losses were still high because of the close crowding together of conductors in the same circuit and of the fact that even the best insulation which could be provided, namely dry paper, resulted in considerably more energy losses to the telephone conversations than take place on open-wire circuits in which the conductors are separated at very considerable distances by air.

The use of repeaters in cable circuits made it possible to get high net efficiency over long distances using small conductors, since it was possible to compensate for the relatively large loss by the repeated gains introduced into the circuit by repeaters suitably located about 40 to 50 miles apart. That this might be done, however, required a reduction of manufacturing tolerance limits for cables, loading coils, and apparatus and care in the design, construction, and maintenance of the cable circuits. Also, new loading systems were designed which transmitted a broader band of frequencies than those transmitted by the earlier systems. This was desirable both because of the improved clearness of speech resulting from the broader band of frequencies itself and also because the use of the new loading systems made it possible to provide for better repeater balance within the band transmitted.

While these improvements made possible a very large extension in the distances over which good transmission could be given on small gauge cable circuits, it was found that, with many repeaters, the balance difficulties were still sufficient to justify the development and use for the longer circuits of a different arrangement. This arrangement, which is shown schematically in Fig. 14, consists of using for each telephone circuit two pairs or two transmission channels, each equipped simply with one-way amplifiers and thus arranged to transmit the telephone currents in one direction only. Two such one-way channels, oppositely directed, are connected together at the terminals of the circuit by apparatus similar to that used for associating amplifiers with two sections of line in the ordinary telephone repeater, including apparatus for balancing the line or the equipment to which the circuit is connected when in use. The complete circuit is thus reduced at its terminals to two wires like any other telephone circuit. From the fact that it uses two channels for transmission in opposite directions, it is called a four-wire circuit. These channels may, however, be either side or phantom circuits as in the case of an ordinary repeatered telephone circuit.

With the four-wire circuit, since the two directions of transmission are kept isolated from each other throughout, there is no need for providing balance except at the terminals of the circuit and this makes possible the use of higher repeater gains and, therefore, a higher net efficiency of transmission with such circuits for long distances than would be possible with the other form of circuit (generally called two-wire circuit). Circuits of this four-wire type are now generally used in toll cables for distances more than about 100 to 150 miles.

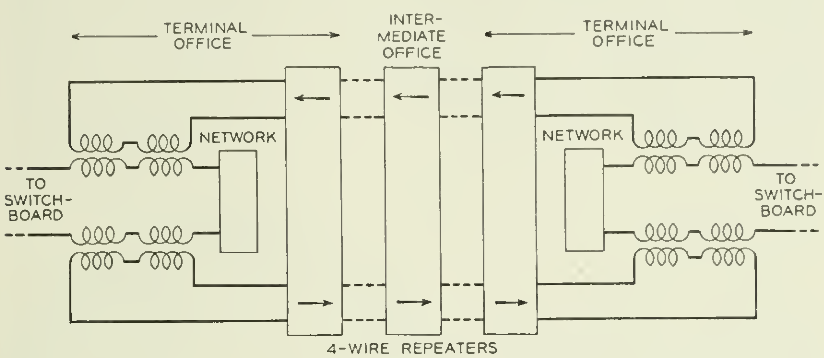


Fig. 14—Schematic of a four-wire circuit using two one-way transmission paths.

Mention was made in the first chapter of this statement of the refinements in manufacture and in installation procedures which were necessary in order to produce suitable interurban toll cables, particularly cables designed for the use of phantom circuits. With the extension of toll cables to great distances equipped at frequent intervals with telephone repeaters, additional refinement in design and in construction was necessary in order to prevent crosstalk between the different circuits in the cable. This includes the physical separation in different parts of the cable of conductors used by four-wire circuits for the opposite directions of transmission. Even with all the refinements which have been worked out, crosstalk remains today one of the major factors to be considered in the engineering of long telephone circuits.

The velocity of propagation of telephone currents over circuits is high so that in all of the early telephone development the length of time required for propagation over the longest circuits used was not sufficient to introduce any new difficulties in the problem of providing good telephone transmission. The velocity of transmission, however, varies with the type of circuit, is lower on loaded circuits than on non-loaded circuits, and in loaded circuits in cable in common use is as low

as 10,000 miles a second. With the greater distances for which cable circuits of the four-wire type could be used, it was found that the time of transmission required at a velocity of 10,000 miles a second was great enough to introduce additional difficulties in the provision of satisfactory transmission. The nature of these difficulties and of the means adopted for overcoming them will be discussed a little later. However, it must be mentioned here that to overcome these difficulties, it was necessary, for these longer circuits, to devise new loading systems which provided circuits with a velocity of 20,000 miles a second and at the same time had the advantage of transmitting a broader band of frequencies, although they had the disadvantage that the circuit had higher transmission losses per mile and therefore required greater amounts of amplification. These higher velocity circuits are in general use for all long cable circuits and have been found satisfactory up to the greatest distances spanned by cables in this country at the present time, namely, approximately 2500 miles.

Cables are placed either underground or supported overhead from a steel messenger strand strung on poles. At the present time approximately 47 per cent. is overhead and 53 per cent. underground. For the most part the underground cable is pulled into permanent underground conduit of vitrified clay. Some use has been made, however, of cable buried directly in the ground, the lead sheath being protected either by layers of jute impregnated with asphaltum compounds or by a combination of such layers of jute and wrappings of steel tape. A small use has also been made of a single duct made of compressed fibre for the protection of underground cables.

The conductors used for long telephone circuits are quadded for phantom operation and are largely of 19 A.W.G. although some use has been made of 16 A.W.G. for the shorter circuits because of a possible saving in the numbers of repeaters with the larger gauge in those cases. Many of the cables include a number of special 16-gauge pairs provided specifically for program transmission circuits and equipped with loading and amplifiers designed particularly for that form of service. Figure 15 shows schematically the arrangement of conductors of a standard type of full size cable (outside diameter 2⅝ inches) which is in common use.

With these developments and other auxiliary developments which will be discussed later toll cables have come to have a very important place in the provision of toll telephone service by the Bell System. The percentage of toll wire in cable has increased from 30 per cent. in 1915 to 82 per cent. at the present time. The present toll cable net-

work is indicated in Fig. 16. It will be noted that this network connects together almost all of the major places between the Atlantic Seaboard on the East; Atlanta, Georgia and Dallas, Texas on the South; Western Texas, Kansas City, and Omaha on the West; and Toronto, Montreal, and Bangor on the North. In addition, there are other sections of toll cable connecting important centers as San Francisco–Los Angeles and Miami–Palm Beach. These cable systems provide a storm-proof outlet for telephone circuits to 155 out of a total of 210 cities over 50,000 population in the United States and Canada, and cover the major part of the United States in which open-wire lines are subject to interruption by severe sleet storms. The cable network includes at the present time about 27,000 miles of cable and 12,500,000 miles of conductor.

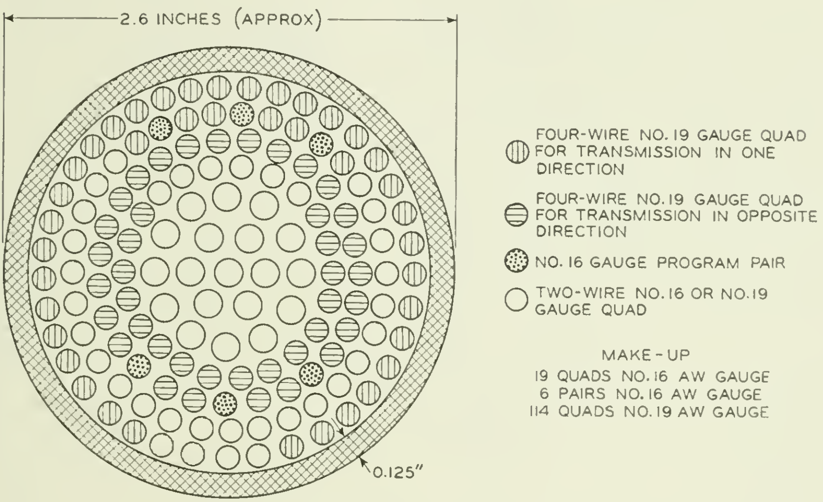


Fig. 15—Cross-section of typical toll cable.

The Time Factor In Telephone Transmission

In the above discussion of toll cable systems it was mentioned that it became desirable for the long circuits in cable to provide a type of circuit having a higher velocity of transmission than that of the loaded cable circuits previously in use. The effects of the length of time required for transmission over long circuits, while particularly noticeable in long cable circuits, are of importance in long open-wire circuits as well. These effects are briefly discussed below.

On non-loaded lines, either in open wire or in cable, the velocity of transmission of telephone currents over the line conductors is high,

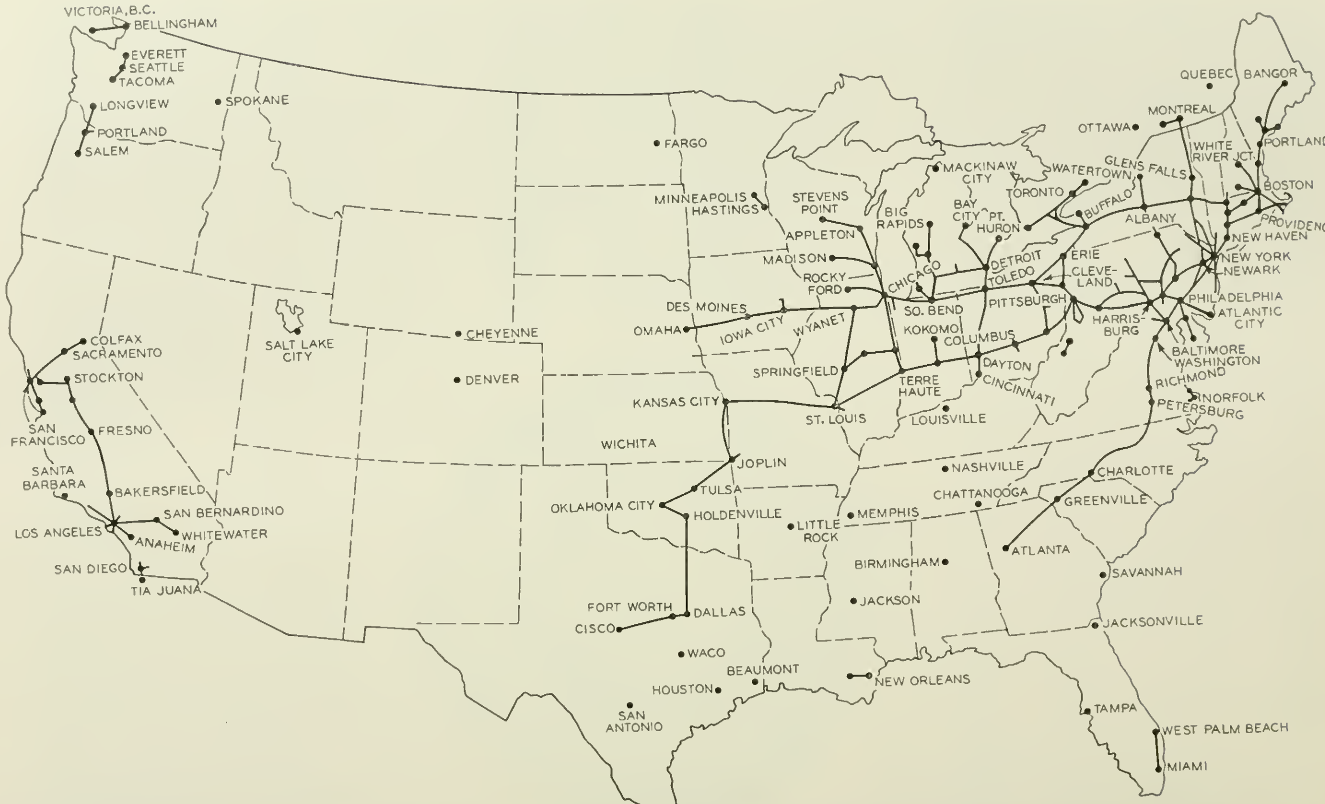


Fig. 16—Main toll cable routes—Bell System.

approaching as an upper limit the velocity of propagation of light, namely, 186,000 miles a second. On loaded circuits this velocity is greatly reduced. Loaded toll cable circuits in common use over moderate distances, as already mentioned, have velocities of 10,000 miles a second and still lower velocities were associated with some of the earlier types of loaded toll cable circuit.

The transmission of telephone currents over long circuits is accompanied by reflections of a part of the current at points where the electrical characteristics of the circuit change, particularly at the terminals where it is not practicable to get a close match between the characteristics of the toll circuit and of the various local circuits and terminal equipments to which it must be connected. Considering for the moment only this terminal reflection, speech over the circuit is not only transmitted directly but also a part of the transmission current is reflected back and forth between the terminals producing delayed sounds analogous to the echoes produced when one talks in the face of a distant cliff or building. For example, over a 1000-mile circuit with a velocity of transmission of 10,000 miles a second, the time required for transmission in one direction is .1 of a second and for a round trip of the circuit .2 of a second. On such a circuit the talker may hear in his receiver the echo of his own words .2 of a second after they are spoken and the listener at the other end may hear not only the direct transmission but an echo delayed by .2 of a second. Such effects, if sufficiently great, seriously interfere with the conversation, the amount of interference increasing with the amount of delay.

The reduction in the effect of echoes is partly taken care of by improvements in the design, both increasing the velocity of transmission over the circuit and reducing the amount of reflected current. In addition, special devices known as "echo suppressors" are used to reduce further the effect of echoes. In the echo suppressor a small part of the voice current is amplified and rectified and used to control the circuit in such a way that during the conversation the circuit is operative only in one direction at a time, the interruption of the return path serving to prevent the transmission of echoes. As people speak alternately from the two ends of the circuit, this control is automatically shifted so that the words will be fully transmitted in each case.

While echo suppressors are very successful and are widely used, they have certain limitations and, in spite of their use, echoes remain an important factor to be considered when engineering and laying out long telephone circuits.

Another effect of the time of transmission arises from the fact that, generally speaking, the components of different frequencies making

up the voice currents are not all transmitted with the same velocity over the circuit. Often the frequencies in the middle of the range 1,000 to 1,500 cycles arrive first and the highest and lowest frequencies arrive somewhat later. This difference is inappreciable on short circuits but for the longest circuits, if not corrected for by suitable design, may become great enough to be appreciable. Under those conditions a distortion of the speech takes place which interferes with the ease of understanding and, in extreme cases, may seriously impair transmission.

This type of effect can be compensated for by the installation at intervals along the circuit of networks designed to introduce additional delay in the transmission of the frequencies in the middle of the range so that all frequencies will arrive at the distant end more nearly at the same time. Up to the present time, the improved design of circuits used for the very long distances has sufficiently kept down the amount of this distortion so that special compensating arrangements are not necessary to message circuits but they are commonly used in circuits for some of the special services, where transmission requirements are more severe.

Still a third effect of the finite velocity of transmission over telephone circuits is to be found in the time of transmission itself. In the ordinary case, the elapsed time between the speaking of a word at one end of the circuit and its reproduction at the distant end is inappreciable but for very long circuits this requires consideration. Telephone conversations, like face-to-face conversations, involve the repeated interchange of information. Even if one person is doing the talking, he receives frequent acknowledgments from the other that he is followed and understood and, in the case of telephone conversations, those acknowledgments must be vocal in character. If too great a time is required for the transmission of the speech and the return transmission of the acknowledgment or replies, the vocal interchange of ideas is interfered with.

These considerations have led to the preliminary conclusion that the total time of transmission over any telephone circuit should not exceed about ¼ of a second. It would mean that the velocity of transmission 20,000 miles a second now used for long toll cable circuits would not be adequate at some future time for connections between widely separated parts of the earth's surface. Fortunately, the trend of development of very long circuits is for various reasons in the direction of higher velocity circuits, as will be made apparent in the next section, so that it is anticipated that this limitation, except in perhaps a few special cases, will not be difficult to overcome.

Multi-Channel Telephone Systems

It was pointed out under "Early Developments" that, for clear transmission, telephone circuits must transmit a band of frequencies, the minimum band used for new telephone circuits being between approximately 250 and 2,750 cycles. However, many telephone lines can be made suitable for transmitting a much broader band of frequencies, namely, frequencies running up into the tens of thousands or, by applying the latest developments, to hundreds of thousands of cycles. This fact naturally raised the question whether some means could not be devised for operating a multiplicity of telephone channels on one circuit using this broader frequency range.

The general idea is as old as telephony itself or older as applied to telegraphy. Alexander Graham Bell's invention of the telephone came, in part at least, through his experimentation in means of providing several telegraph channels over one circuit by using currents of different frequencies. The fundamental principles of multiplex telephony were early thought of and well understood. They involve:

(1) Means for so varying a high-frequency current (called a "carrier") that, with this variation, it represents the sounds to be transmitted over the telephone circuit just as do the voice currents produced by the telephone transmitter in the range 250 to 2,750 cycles. As ordinarily carried out, this involves the control of the amplitude of the carrier current in proportion to the instantaneous values of the voice-frequency telephone currents, this process being known as "modulation."

(2) Correspondingly, means for reproducing the sounds transmitted by suitably operating upon the modulated high-frequency current. This is done by reproducing from this current the ordinary voice current (a process known as "demodulation") and applying this voice current to an ordinary telephone receiver.

(3) Means for joining the modulated carriers of different frequencies so that they can be transmitted over the same telephone wires, and for completely separating them from each other at the receiving end by virtue of their different frequency ranges so that each modulated carrier can be demodulated in a separate receiving circuit and the various conversations carried on simultaneously without interference. This function has been termed selectivity.

While the fundamental ideas as outlined above are old, the physical means by which successful carrier current telephony could be made practicable did not become available until the period 1913 to 1918. In that period, the successful development of the vacuum tube for use in telephone repeaters produced a device which, with different circuit arrangements, could be used satisfactorily for generating carrier currents, modulating them with telephone currents, and for reproducing the telephone currents from the modulated carrier currents. At about the same time, marked advances were made in the development of means for separating into any desired groups a mixture of currents of different frequencies transmitted over the same conductors. These means may be considered in principle an elaboration of the elementary apparatus of this type, called "composite sets," long in use for separating telephone and telegraph currents transmitted over the same circuit by reason of their difference in frequency. The more complete solution of this general problem was made by the development by the Bell System of the "electrical filter."

With these new tools it became possible to develop carrier telephone systems suitable for commercial service. Such systems were first introduced into the plant of the Bell System in 1918. Since that time their use has spread widely, particularly over non-loaded open-wire circuits of the System.

The most important type of carrier telephone system in general use is the Type C. One terminal of such a system is indicated schematically in Fig. 17. With this system three carrier channels, marked *A*, *B*, and *C*, and a voice-frequency channel, marked *V*, are transmitted simultaneously over one pair of wires. The four circuits as they appear at the toll switchboard are alike and are treated indiscriminately by the operator. Coming from the switchboard as indicated at the left of Fig. 17, the three carrier channels first pass through three individual sets of carrier equipment. In each of these sets of carrier equipment, the circuit is separated into transmitting and receiving paths. The transmitting path is passed through a modulator in which the voice currents received from the switchboard act upon a carrier and produce modulated carrier currents, and through an electric filter to the general transmitting circuit indicated on the drawing. The receiving channel is connected to the general receiving circuit through an electric filter and through a demodulator by means of which the received currents are caused to reproduce voice-frequency currents similar to those delivered to the circuit at the other end.

As the next step the three transmitting channels are brought together through a common amplifier and transmitted through a "band"

filter to the line filter. Similarly, the three receiving circuits are brought together by a common receiving amplifier with which is associated a "band" filter, which in turn is connected to the line filter. The line filter consists of two parts, one of which permits the carrier currents of Channels A, B, and C, to pass but excludes the voice-frequency currents and through this operation of the line filter the carrier currents are transmitted to the line. The voice-frequency circuit V is connected to the line through the other part of the line filter which permits voice-frequency currents to pass but excludes all of the carrier-frequency currents. These currents of different frequencies from four channels are then transmitted together over the line and at the receiving end are separated by apparatus similar to that indicated in this sketch.

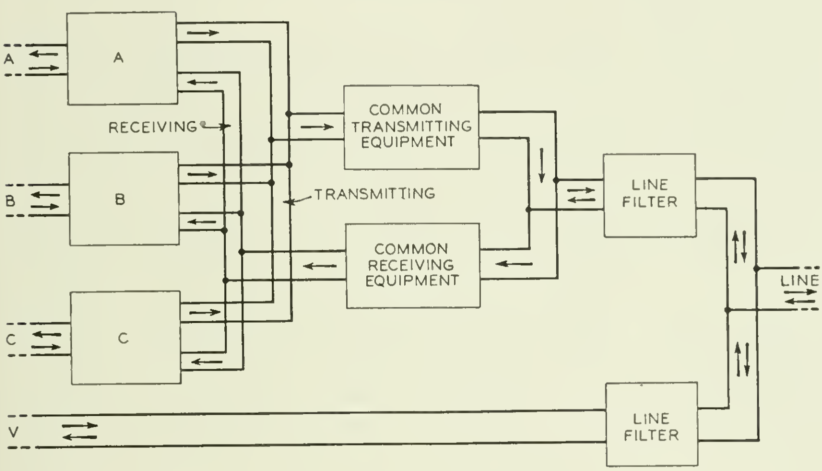


Fig. 17—Schematic arrangement showing the association of a type "C" carrier telephone terminal with the telephone line.

At intermediate points along the circuit, it is necessary to install amplifiers for the carrier-frequency currents as well as for the voice-frequency circuits. At these points, the carrier-frequency currents are separated as a whole from the voice-frequency circuit, a telephone repeater being used for the voice-frequency circuit and a carrier repeater consisting of two amplifiers with electrical filters to separate the two directions of transmission, being used to amplify the three carrier circuits as a group. After amplification, the carrier and voice-frequency currents are again brought together for transmission over another section of line.

It is to be noted that the carrier circuits, like the four-wire cable circuits, use different channels of transmission in the two directions, thus avoiding the difficulties of balance which would otherwise be encountered. In the case of the carrier systems now in general use, however, the two channels are channels of different frequencies operating in opposite directions on the same pair of wires rather than using two separate pairs of wires. As a result the Type C carrier system providing three two-way telephone channels transmits over the line six bands of carrier frequencies, one for each direction for each of the channels. The maximum frequency used by the Type C system is about 30,000 cycles.

In addition to the Type C system just described, there is used in the Bell System a simpler single-circuit carrier system (Type D) providing one carrier circuit in addition to the voice-frequency telephone circuit.

In the case of carrier systems as in the case of telephone repeaters, their application to the telephone plant involved not only the development of the system itself but the development and application of new practices to the telephone plant. This came about from the fact that the plant, heretofore designed primarily for the transmission of ordinary voice frequencies, that is, currents up to about 3,000 cycles per second in frequency, was now called upon to transmit currents up to 30,000 cycles successfully and without interference. In order to do this, it was necessary in the open-wire circuits to use non-loaded pairs and where loading was necessary in short sections of incidental cable in such circuits, to design new loading systems with loading coils of small inductance placed at frequent intervals which would transmit these higher frequency currents. A major problem of adapting the plant to the use of these currents arose from the increasing tendency with higher frequencies for currents flowing in one circuit to induce currents into other circuits in the vicinity. The transposition systems used to prevent crosstalk between voice-frequency telephone circuits on the same pole line were wholly inadequate to prevent crosstalk of the carrier currents and without extensive changes such crosstalk would have been far too great to make possible the satisfactory use of carrier systems. New systems of transpositions involving a large increase in the number of transpositions used in a given section of line were designed for this purpose. Also, it was found that for the largest use of carrier systems it would be necessary to give up the use of phantoms on the circuits involved and also to rearrange the conductors to provide less space between the two wires of the pair and greater amounts of space between the pairs on the same crossarm.

These new construction arrangements have been worked out and applied where the extensive use of carrier is sufficiently important to justify them.

As a result of these various developments, an extensive use of carrier systems has been made in the Bell System plant. This is indicated in Fig. 18, which shows the routes on which carrier systems are used at the present time. The total circuit mileage in service provided by carrier systems is about 400,000 miles, which is over 8 per cent. of the total toll circuit mileage in service.

Up to the present time, the applications of carrier have been confined to open wire, including relatively short sections of incidental cable in the open-wire circuit. Further advances in the art, particularly in the design of very stable amplifiers capable of amplifying simultaneously a large number of carrier channels of different frequencies without mutual interference and improvements in the design of electrical filters to make them less expensive and more effective have opened the way for broader applications of carrier. These broader applications include the prospective use of carrier on non-loaded cable circuits with amplifiers spaced at intervals of twenty miles or less. Systems are now being developed for this service by which it is expected to get 12 one-way channels on a single non-loaded cable pair, and with cables of special construction, such as the coaxial cable, on which experiments are now being made, several hundred one-way transmissions may be obtained on a single unit.

In view of these developments under way, and further prospective improvements in carrier systems applicable to open-wire circuits, it is evident that this form of transmission will have in the future a rapidly growing field of use in the telephone plant.

Associated Technical Developments

The successful operation in a practical telephone plant of the new types of circuit for transmission over very long distances both in cable and in open wire required, in addition to the main developments briefly outlined above, the developments of a number of associated technical arrangements. Some of the more important of these are briefly outlined in the following paragraphs.

Regulators

In the long telephone circuits made possible by the use of repeaters, having a number of repeaters at different points along the circuit, the net transmission efficiency of the circuit is the result obtained by balancing the amplification of telephone currents in the repeaters

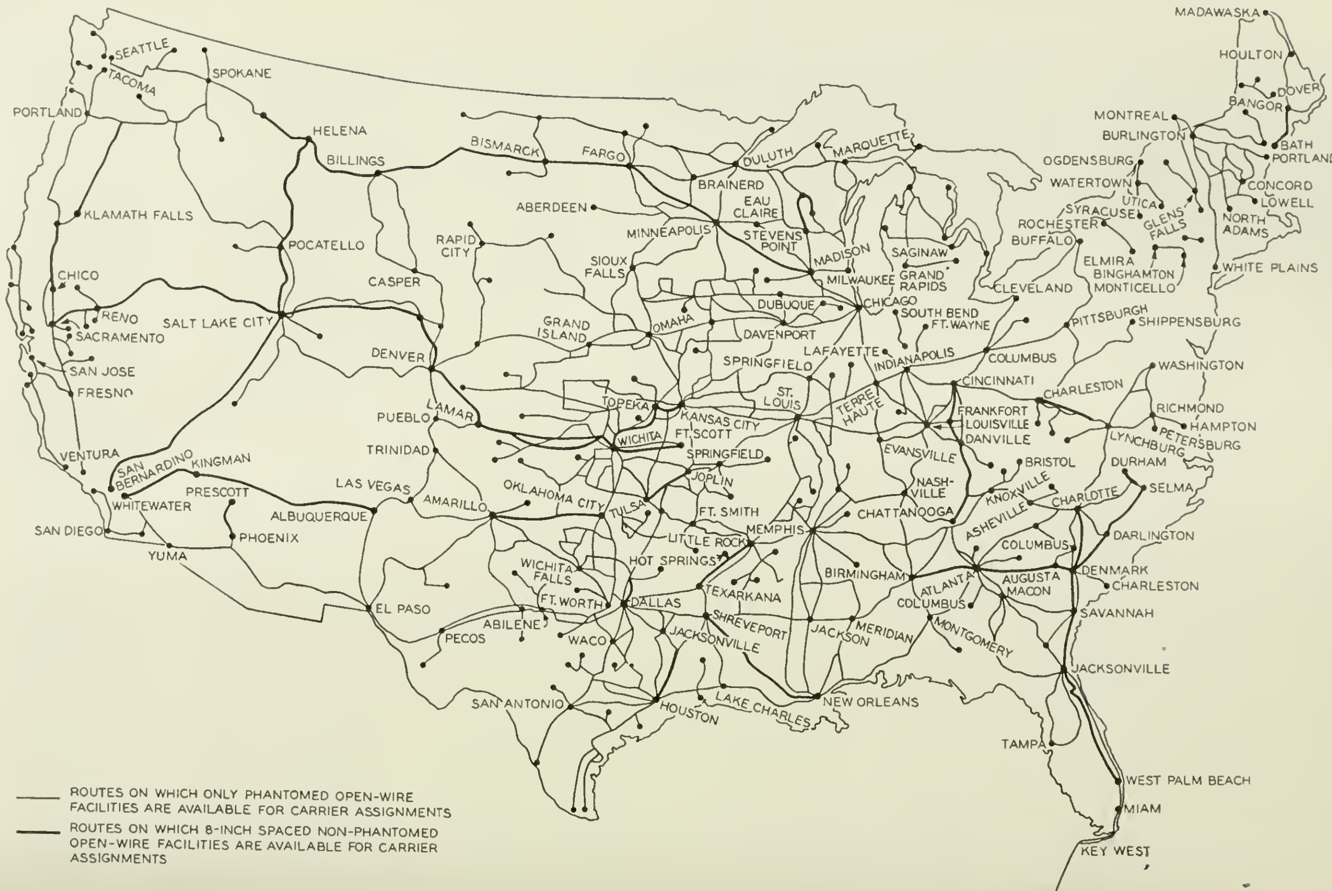


Fig. 18—Routes of the American Telephone and Telegraph Company and Associated Companies in the United States on which carrier telephone facilities are available.

against the attenuation of the telephone current in the various line sections of the circuit. As a result, with increase in length such circuits become increasingly susceptible to the effect of variations in circuit efficiency caused by changes in weather conditions. In the case of toll cable circuits this variation is primarily due to changes in temperature and in a long cable circuit such changes may in a single day make a 10,000 fold difference in the overall efficiency of the cable. In the case of open wire circuits the change is mostly due to rain and is particularly prominent in the highest frequency carrier systems.

In order to offset these variations and provide circuits of approximately constant overall net efficiency these longer circuits are equipped with regulating systems which make it possible to offset the variations in the efficiency of the circuits by either manually or automatically changing the gain of certain of the repeaters.

In the case of the cable circuits the regulation system, known as a "pilot wire regulator," is automatic. It makes use of a direct-current channel over a metallic composite circuit, variations in the resistance of this pilot wire which result from the variations in temperature causing automatic adjustment of the regulating repeaters. A single pilot wire with its associated regulating equipment can be used for controlling all of the circuits in a cable. Sometimes as many as 300 circuits or more in number are regulated with a single pilot wire. In order to get sufficiently accurate regulation and because of practical layout considerations the cable is generally regulated separately in sections of 100 to 150 miles in length.

In the case of a type C carrier system on open wire the regulating system depends upon attenuations of carrier currents transmitted over the same pair as the carrier system. On most of the type "C" systems the regulators give an indication of the net efficiency, which is kept within a prescribed limit by manual adjustment of the repeaters. In some cases apparatus providing this adjustment automatically is in use.

Equalizers

The transmission efficiency of telephone lines is generally different for the different single frequencies comprising the voice-frequency band. This variation is particularly great on some loaded cable circuits where the maximum frequencies transmitted approach the maximum frequencies which the circuit is capable of transmitting. If such variation in efficiency were permitted, resulting in the higher frequencies having much greater losses than the lower frequencies, the normal relative proportion of different frequencies would in some cases be so distorted that the transmitted speech would not be clear or

even could not be recognized. In order to prevent such distortion, therefore, it was necessary to develop and apply means for compensating for the variation in the line by variations in the opposite direction. These means are called attenuation equalizers. In the telephone message circuit these attenuation equalizers are generally designed to be an integral part of the telephone repeaters themselves.

Signaling Systems

With the exception of a few special cases all telephone circuits for message telephone service are provided with a means by which signals can be transmitted from one end to the other in order to call to the circuit an operator (or sometimes in the case of dial systems, a machine) or subscriber. In the relatively short circuits used for local telephone service this signal commonly is provided either by the flow of direct current which is used to light a lamp or operate a relay or by the flow of 20-cycle alternating current which is used to ring a bell or operate a relay.

Signals based upon the use of direct current are generally used in cases where it is desired to have the signal continue to indicate the condition of a connection throughout the conversation since the direct current can continue to flow over the circuit simultaneously with the voice current during conversation without interference. The most extensive use of this type of signaling is for the shorter circuits. The method of operation of long toll circuits generally is such that no signal is required over the toll circuit during the conversation period but only before and after the conversation, and for such signals alternating current is generally used.

In the early days the alternating currents used for signals over toll circuits were generally the same as those used for signaling over local circuits, namely, 16 to 20 cycles per second. As the simultaneous use of toll circuits for telephone and telegraph expanded, this was modified because of interference which would occur between the telegraph currents and the 16 or 20-cycle signaling currents. For such cases signaling was accomplished by alternating currents of 135 cycles, sufficiently high to avoid interference with the telegraph systems then in use.

With the further development of long toll circuits having many repeaters at intermediate points and also with the development of carrier telephone systems, the satisfactory transmission of 135-cycle current from one end of the circuit to the other became more difficult. At the same time the advance in the art made possible the development of satisfactory and economical signaling systems using interrupted currents of 1,000-cycle frequency. With such a system the signaling

current uses the same transmission path as the voice currents but the system is designed to discriminate between the voice and signaling current. The signaling current is transmitted from end to end over the circuit without the use of intermediate ringers which had been resorted to on the 135-cycle ringing system. This 1,000-cycle system has extended rapidly in its field of use and is now used for the majority of circuits over 100–150 miles in length.

Telegraph System

The early methods by which provision was made for the simultaneous use of telephone circuits for telephone and telegraph service were described in the section on "Early Developments." With the general extension in the telephone plant of the improved types of telephone circuits described above, and with the growth in extent and requirements of the private line telegraph service, it became important to devise new types of telegraph circuits adaptable for use with the new types of telephone circuit and suitable to meet the increased telegraph requirements.

One such new form of telegraph circuit was the metallic telegraph system designed for use on telephone toll cables simultaneously with the use of the same conductors for telephone service. In order to avoid interference between the telephone and telegraph circuits it was necessary to use relatively low voltages and currents on the telegraph circuit. With these low voltages and currents grounded telegraph circuits were impracticable because of outside interference and it was necessary to use metallic circuits in which no use was made of the ground for the transmission path. With the new metallic circuits voltages of 34 volts and currents of 4 milliamperes were used compared with 130 volts and 60 milliamperes in the grounded direct-current telegraph systems.

Another form of telegraph circuit which was developed for use over telephone toll cables is the so-called voice-frequency telegraph system. With this system by carrier current methods the telephone channel is split up into 12 telegraph channels, each of which is suitable for use as an independent telegraph circuit, one telephone circuit thus providing 12 telegraph circuits. In this case, the telephone circuit cannot be used simultaneously for telephone and telegraph circuits. This system is designed to be applied either to a four-wire cable circuit, or to a carrier telephone circuit which, like the four-wire cable circuit, consists of two channels for transmission in the opposite directions. It has large advantages for long circuits because of the fact that no apparatus is required at points intermediate between the terminals,

including points where cable and open wire are joined, other than the standard telephone repeaters and associated apparatus already provided with the circuit for telephone purposes.

Still a third type of telegraph system devised to meet the new conditions is that known as the high-frequency carrier telegraph system. This system is designed for application to open wires. It applies to the provision of telegraph circuits the same principles as are applied in the carrier telephone system for the provision of telephone circuits. It uses frequencies above the voice range, roughly in the range from 3,000 to 10,000 cycles, thus permitting the continued use of the conductors for a voice-frequency telephone circuit simultaneously with its use for high-frequency carrier telegraph circuits. With this system 10 two-way telegraph circuits are provided.

Auxiliary Apparatus and Equipment

In addition to the main items described above, developments of other apparatus and equipment auxiliary to the telephone toll circuits were made necessary by the general use of repeaters and carrier systems. Power plants providing current to the filaments and plate circuits of the vacuum tubes used in repeaters and carrier systems had to be provided having much closer voltage regulation than had heretofore been necessary for the earlier types of telephone equipment. New forms of testboards were required and new types of arrangements of distributing frames and of protective apparatus. Plans were developed for the economical arrangement of the new types of equipment in large offices. All of these things while essential for the proper operation of modern toll telephone circuits probably do not need detailed discussion in this statement.

Another type of equipment which had to be developed was that for carrying out the various forms of electrical test necessary to assure the proper operation of these new telephone circuits. The development of this equipment and of the new maintenance methods which made use of this equipment is of sufficient general importance so that it is briefly discussed in the next section of this statement.

Development Of Methods Of Measurement And Maintenance For Toll Circuits

The history of toll service has been a story of the continuous application of new scientific instrumentalities. The laboratory experiments of one day become the regular service-giving apparatus of ever-growing complexity. Maintaining this complicated equipment at a high state of efficiency has been accomplished through methods of

measurements, developed either to make possible the measurement of electrical quantities for which no methods of measurement existed previously or else to make possible measurements in large numbers on a routine basis at little expense which previously were delicate, expensive and confined to the laboratories. It would be out of place to include in this report a general discussion of the development of these methods. Mention will be made, however, of certain items which have particular reference to the new types of toll circuit discussed above.

An important tool in electrical measurements is the Wheatstone bridge devised originally for the accurate measurement of resistances. In dealing with telephone circuits where the performance of circuits and apparatus in the transmission of alternating currents is important, it was necessary to expand the Wheatstone bridge for alternating current use. This involved providing elements for the Wheatstone bridge having not only a known resistance to the flow of direct currents but also a known resistance and reactance to the flow of alternating currents of the frequencies at which measurements were to be made. It was soon found, however, that with frequencies as high as those required in telephone measurements, running up to two or three thousand cycles, the resistance and reactance of the elements of the Wheatstone bridge were not well known and varied depending upon the number of elements connected in the circuit. This variation was due to the effect of incidental capacitances between the elements of the bridge and between them and the ground. In order to overcome these difficulties, G. A. Campbell devised an arrangement of shields by which variation in the effect of these incidental capacitances was prevented and in this way produced bridges for alternating current use which would give accurate results over the range of frequencies required in telephonic measurements.

An interesting example of the application of the shielded impedance bridge to practical telephone measurements is presented by what is called the "capacity unbalance (testing) set." This testing set is designed to measure the very small capacitances between individual wires of a short section of toll cable or more specifically differences between these capacitances for the individual wires of two pairs or of two quads, expressing these differences in such a way that they are directly proportional to the contribution made by the capacitances in the short section of cable to crosstalk between circuits using the pairs or quads thus tested. The purpose of the test is to give information to the splicing forces, which, properly interpreted by them, enables them to splice together pairs and quads in adjacent lengths in such

combinations that the crosstalk unbalances in adjacent sections tend to neutralize each other and the crosstalk between all pairs and quads in the cable when completed will be small. The capacitance differences measured in individual lengths are only a few millionths of a microfarad. This measurement, originally possible only under carefully controlled conditions, is, by the use of the capacity unbalance testing set, reduced to a routine part of the work of the construction and cable splicing forces.

Another interesting kind of measurement bearing in a very important way on the maintenance of the efficiency of telephone toll circuits is measurements of transmission efficiency, that is, of the power output of the telephone circuit in proportion to the power input of alternating current at the distant end. In order that such a measurement may represent the efficiency of the circuit for the transmission of telephone currents, it is necessary not only that the frequency of the testing current correspond to one of the important frequencies of telephone currents (1,000 cycles is ordinarily used when only one frequency is necessary) but also that the amount of power transmitted correspond approximately to the average power of telephone currents. For this reason the standard input power for such tests is one milliwatt and the power received at the other end of the telephone circuit is often one-tenth of that or less.

When tests of this sort were first made as a part of the routine work of maintaining telephone toll circuits the only available instrument sufficiently sensitive to measure the received power and practical for use under field conditions was the combination of the telephone receiver and the ear. In making such a measurement power was transmitted alternately over the circuit to be tested and over an artificial circuit whose efficiency was known and adjustable, the adjustment being made until the received sound was equally loud in the two cases. Then with the further development of the art sensitive receiving instruments became available which were substituted for the telephone receiver and the ear, the adjustment then being made of the artificial telphone line until the received power as indicated by the sensitive meter was equal to that received over the circuit under test. The perfection of instruments of sufficient sensitivity for this measurement and yet sufficiently rugged to be practicable for use by the regular telephone maintenance forces constituted a great advance in the development of measuring systems for telephone transmission. The most satisfactory instruments of this type made use of vacuum tubes to provide the necessary sensitiveness.

A still further improvement has been made by development of instruments which within a limited range showed directly by their amplitude of deflection the amount of power loss in the telephone circuit. With this latter development the artificial circuit is entirely dispensed with, the standard amount of alternating current power applied to the circuit at one end and the meter connected to the other end. These great advances in the technique of measuring instruments provided an ease of measurement almost comparable to the ease of the measurements commonly made in power transmission systems where the large amounts of power available made the development of satisfactory instruments very much less difficult. Now a still further advance in these methods of measurement has been made by devising arrangements such that the deflection of the instrument is indicated in an enlarged scale on an illuminated screen. This makes it unnecessary to transport the instrument to the terminal of the circuit and makes it possible in a repeater office, by making connections in one part of the room so that the circuit is connected to the receiving instrument, for the maintenance man to read the deflection of the meter at a distance thus further cutting down the time required for tests of this nature.

Other types of tests on telephone toll circuits for which special measuring apparatus and measuring methods have been devised include measurements of the crosstalk between circuits, measurements of the noise currents induced in circuits by other electrical circuits, such as electric power circuits, measurements of the uniformity of electrical impedance from the standpoint of suitability for operation with repeaters, measurements of the amplification of telephone repeaters and measurements of the thermionic activity of the vacuum tubes.

While the above discussion refers to instruments for the measurement of alternating currents in what is called the voice-frequency range, that is, up to about 3,000 cycles per second, the introduction of carrier telephone and telegraph systems made necessary the development of similar measuring instruments for the higher frequency currents used in carrier, namely, up to about 30,000 cycles per second. With the expected use in the future of currents up to frequencies of 100,000 or 1,000,000 cycles or more the range of field measuring apparatus will, of course, have to be greatly increased.

The use of these special types of apparatus for making necessary electrical measurements has required a large amount of instruction of the maintenance forces. Also, it was necessary to devise systems of test and adjustment using these measuring methods by means of

which the transmission performance of toll circuits of the new types can best be maintained at the desired standards. This involves a determination of the kind of tests and the limits of adjustment necessary for the different types of apparatus included in these circuits, the frequency of tests and the desirable range of performance results for the maintenance of a high quality of service over these circuits, with the least practicable expense for their maintenance. Maintenance routines of this sort are developed from time to time with each new type of circuit and amended to accord with modifications in the details of the circuits or to take advantage of the results of field experience.

Special Services

With a nation-wide network of poles, wires and circuits available for telephone message purposes, and with its accumulated knowledge concerning technical communication problems the Bell System, as the demand has arisen, has naturally been in a position to analyze the technical requirements of the special communication services and to provide suitable facilities for them. The earliest demand for intercity circuits for special services were for private telephone circuits between telephones in different cities and for private line telegraph circuits. Since that time developments in the communication art, such as radio broadcasting and the transmission of pictures over wires, have created additional demands.

The toll wire plant of the Bell System can be used either interchangeably or simultaneously for telephone message service and many of the special services. In addition, the telephone message service and practically all the special services make common use of many other parts of the toll plant, such as poles, conduits, buildings and power plants.

Some of the special services which make use of telephone circuits or of circuits similar to telephone circuits involve special requirements for satisfactory transmission. This is best illustrated by the transmission of programs for radio broadcast stations, a service which is given on a nation-wide basis over the toll plant of the Bell System. The principal reason for the wide difference in technical requirements of program transmission circuits and of telephone message circuits is that, unlike the message circuits, program transmission circuits are required to transmit music as well as speech. The satisfactory reception of transmitted music requires the transmission of a broader band of frequencies than is necessary for speech alone. The national program transmission networks of the country at the present time, consistent with the requirements of radio broadcast art, transmit a band of

frequencies of from about 50 to about 5,000 cycles compared with a band of frequencies of 250 to 2,750 cycles commonly transmitted by message circuits, and means by which a broader band of frequencies can be transmitted over program transmission circuits have been developed. Another important requirement of program transmission circuits is that they shall be able to handle a wide range of input power. Generally speaking, the power may be varied over a range of 10,000 to 1, without the overloading of the amplifiers or other apparatus on the circuit at the highest levels or interference with the program by extraneous noises at the lowest levels.

Because of these and other special requirements a large part of the telephone plant devoted to program transmission is designed specifically for that service. In the toll cables special 16-gauge pairs have been placed and these pairs are equipped with loading and with amplifiers designed to produce satisfactory transmission circuits. The equalization for variations in attenuation, the regulating arrangements to assure constant efficiency, and the compensators for the difference in the velocity of transmission of currents of different frequencies present special problems.

On open-wire lines the conductors used are generally of the same type as those provided for telephone message circuits. On the other hand, it is necessary to give up the use of direct current telegraph and generally necessary to give up the use of phantoms on circuits used for program transmission. Also, in some cases the number of carrier channels which can be superposed upon the conductors is reduced. The amplifiers and other equipment used in connection with these conductors for program transmission are of special design.

Not only is the plant for program transmission of special design but even to a greater extent the operating features are special to this type of service. For many conditions continuous monitoring is necessary during the transmission of the program. Special switching arrangements are required to make possible rapid changes in the connection of program transmission networks at the moment of a change in program.

At the present time there are about 60,000 miles of program transmission circuit maintained for full-time and recurring program service, of which about 40,000 miles are in daily service in the Bell System on full-time networks. The extent of the network devoted regularly to this purpose is indicated in Fig. 19.

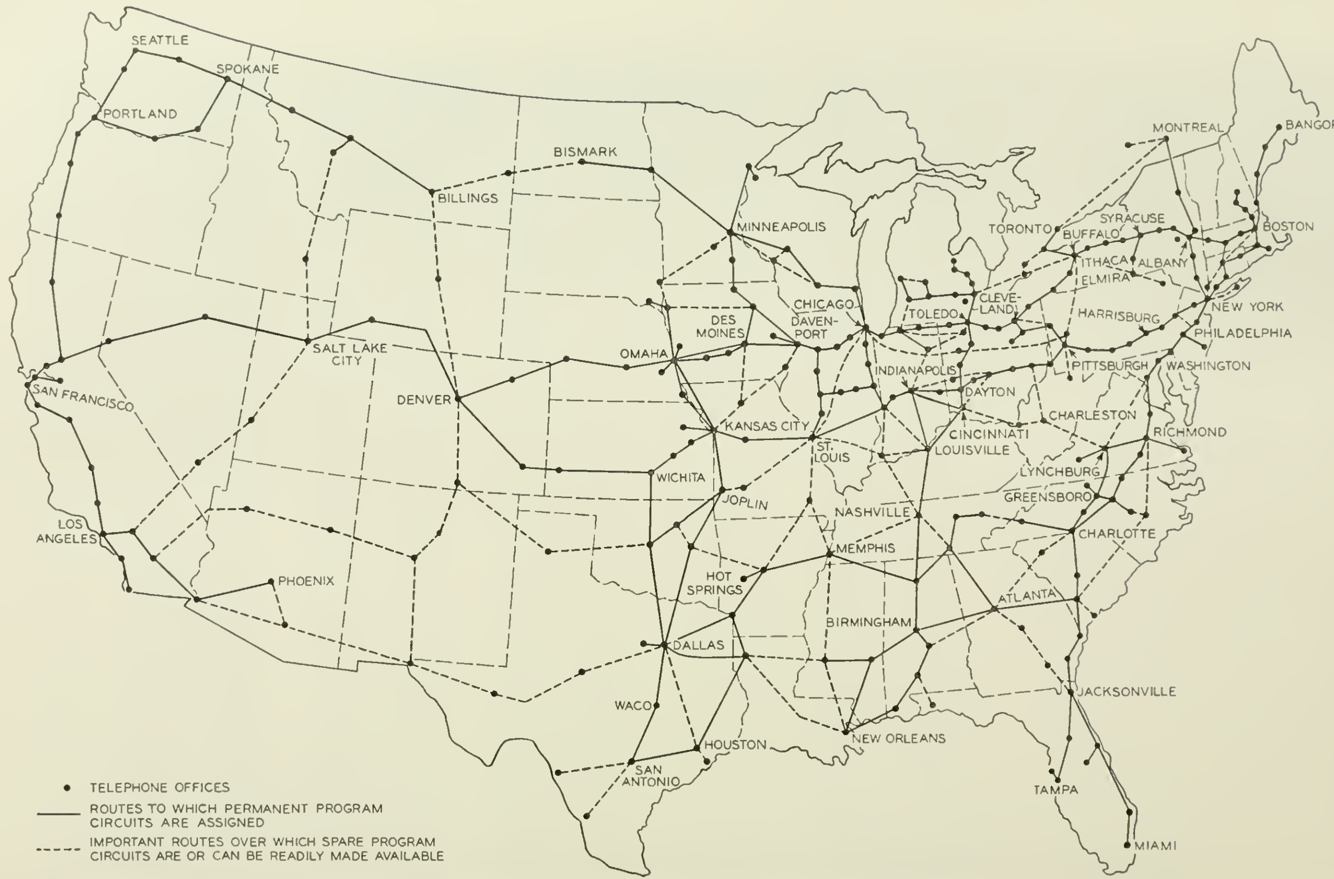


Fig. 19—Routes of the American Telephone and Telegraph Company and Associated Companies in the United States used for radio program transmission.

Telephotography

In 1925 the Bell System inaugurated between a limited number of points a service for the transmission of photographs by wire. Such a system involved the use of telephone circuits, the band of frequencies required for successful transmission being approximately 800–1,800 cycles. This service was discontinued in 1933 because of lack of commercial demand.

At the present time the Bell System is providing to one of the press associations special circuits for their use in transmitting photographs with apparatus owned by them. The type of apparatus used for this circuit is a Bell System development, and represents a marked advance over the earlier apparatus, transmitting pictures at a higher speed and requiring a band of frequencies of approximately 1,200–2,600 cycles. Within the band of frequencies used for the picture transmission a very high degree of equalization of attenuation and velocity of transmission is required. This involves the use of apparatus designed specifically for this service which is associated with regular telephone repeaters. It also requires special attention on the part of the operating forces.

Other Special Services

The Bell System gives an extensive private line telephone service. The requirements for circuits for this service are similar to those for telephone message service and do not require any special discussion.

Also, the toll plant of the Bell System is from time to time used in a limited way for other special services. The private line telegraph service and teletypewriter exchange service are not discussed here, being outside the scope of this statement.

Toll Operating Methods

Operating Method Defined

By "toll operating method" is meant the process by which a toll call is received, recorded, completed and timed. This process is referred to as "handling the call." It relates, for the most part, to the routine and procedure of handling the call, although it must conform to the type of equipment provided, the trunking method involved and the arrangement of the plant. During the development of the telephone business many different toll operating methods have been used, but in the following only the five are described which at various times have come into general use and by which the vast majority of all toll calls have been handled. Such questions as the following are involved in the toll operating method:

How shall the customer's order be received and recorded?

How shall the operator reach the called place?

What combination of plant and method will result in the best service at least cost?

Interrelation of Method and Equipment Design

The operating method and the design of the equipment must be considered together. In many cases the design of the equipment must be changed to permit the use of an improved operating method. In other cases redesign of the equipment is not essential to the improvement of an operating method but in nearly all cases some change in equipment design or arrangement is desirable to permit the best service and the most economical operation of the method. In the normal evolution of the business, improvements in methods and equipment design follow along concurrently. It is not unusual that the greatest amount of work in connection with the improvement of a method has to do with the study and design or redesign of equipment rather than with the study of the operating method alone.

Rearrangement of Plant Brought About by Changes in Method

Various types of switchboard equipment are designed to serve specialized functions in the toll operating room. These various types of equipment must be arranged so that the toll calls can be handled most speedily and economically. Most of the equipment used by operators is provided to make possible the interconnection of a telephone with any other telephone within the exchange or toll network. In addition, however, certain auxiliary equipment is provided which facilitates such interconnection. At information desks no connection is made between telephones but this equipment is provided to make available to operators and subscribers the telephone numbers required in completing connections. In the long distance office the route desk is provided to perform a similar function in connection with the routing of calls. The various operating methods are designed to use these auxiliary equipments to best advantage and the various items of equipment must, therefore, be arranged in such a manner as to meet different operating conditions as methods are changed. Occasionally an improvement in method makes it possible to eliminate one of these auxiliary equipments. An example of this will be shown below in connection with the combined line and recording method. The manner in which various types of equipment are arranged in the operating room, their proximity to each other, their relative locations on different floors of the building, the arrangement of the trunks that

tie them together, the arrangement of the ticket distributing apparatus, all have important effects upon the service rendered by and the efficiency of the operating method. Changes in method, therefore, frequently call for rearrangement of equipment.

Trunking Methods Distinguished from Operating Methods

As soon as the telephone business developed to the point where it became necessary to connect together two telephones not served by the same central office, the arrangements for interconnection between the two offices became an important consideration. Offices are connected together by trunks or toll lines and there must be arrangements for operators to get into communication with each other promptly. In general this is accomplished by signals transmitted over the circuit which later is used for conversation, but sometimes a separate circuit, known as a call-circuit, is used. The manner in which trunks are arranged and used is known as trunking method, as distinguished from operating method which has to do with the manner in which calls are handled. Much of the trunking methods experience gained in handling local traffic has been applied to the handling of toll calls. The more important trunking methods are call-circuits, straightforward, ringdown and dialing. Any of these trunking methods may be used with the various toll operating methods.

Description of Trunking Methods

Call-Circuit Trunking Method

Under this method a call-circuit was provided between the two offices. The terminating end was connected to an operator's receiver and at the originating end, any one of a number of operators could connect her telephone set to this circuit merely by depressing a key. Let us assume, for example, that a call from New York to Philadelphia is being handled by this trunking method. The customer in New York has given the Philadelphia number to the New York operator. The latter depresses a key which connects her telephone set to the call-circuit which at Philadelphia is connected to the receiver of an operator who handles only inward connections from New York. The New York operator listens for a moment, to determine that no one else is speaking on the call circuit, and then passes the Philadelphia number over the circuit. Let us also assume that there are 50 toll circuits between New York and Philadelphia, numbered 1 to 50. At the moment that the New York operator passes the number to the Philadelphia operator, some of these circuits are in use. By glancing at her switchboard the Philadelphia operator determines that circuit No.

13, for example, is not in use. In response to the number passed by the New York operator, she says "one-three." This notifies the New York operator that she is going to connect the required telephone number at Philadelphia to circuit No. 13. The New York operator connects her calling party to this circuit and conversation begins as soon as the Philadelphia subscriber answers his telephone. As on any local call the removal or hanging up of the receiver at the called telephone is indicated to the New York operator by appropriate signal lights.

Straightforward Trunking Method

As improvements in equipment and operating methods were made, the call-circuit trunking method gradually was replaced by the straightforward trunking method. Let us assume that a call is being handled by the straightforward method from office A to office B. The calling party gives the called number to the operator in office A who then makes connection to a trunk to office B. The trunk is connected to apparatus at office B in such a way that when the operator at A makes connection to it she is connected automatically to the receiver of an operator at B and a momentary audible tone indicates to her that the operator at B is ready to receive the call. Upon hearing the tone the operator at A passes the called number to the operator at B who then connects the trunk to the called telephone line. It will be noted that the selection of the trunk or circuit between the calling and the called offices is made by the originating operator under the straightforward trunking method, whereas under the call-circuit trunking method the selection of the trunk is made by the operator at the terminating office.

Ringdown Trunking Method

The ringdown trunking method was the first to come into use and is still used where it is uneconomical to provide the equipment necessary to straightforward or dial operation. Under this method the operator at office A signals the operator at office B by making connection to a trunk or circuit between A and B and by depressing a key which operates a signal associated with the circuit at office B. The operator answers this signal by connecting her telephone set to the circuit and announcing the name of her office. The operator at A then passes the number of the called telephone to the operator at B who makes connection to the called number.

Dial Trunking Method

Under some conditions it is feasible to arrange for the originating toll operator to dial the called number without the assistance of an

inward operator at the called place. By this trunking method the calling party reaches the operator and gives her his call in the usual way. She then makes connection to a trunk or circuit to the called place and upon receipt of the proper automatic signal, indicating that the apparatus at the terminating office is ready to receive the call, she dials the called number. Under this method, as with the call-circuit and straightforward trunking methods, switchboard lamp signals indicate to the originating operator whether the receiver at the called telephone is on or off its hook.

Description of Toll Operating Methods

General Characteristics of Toll Calls and Operating Methods

The percentage of toll calls handled by the various toll operating methods has varied from year to year, as shown in Fig. 20, until at

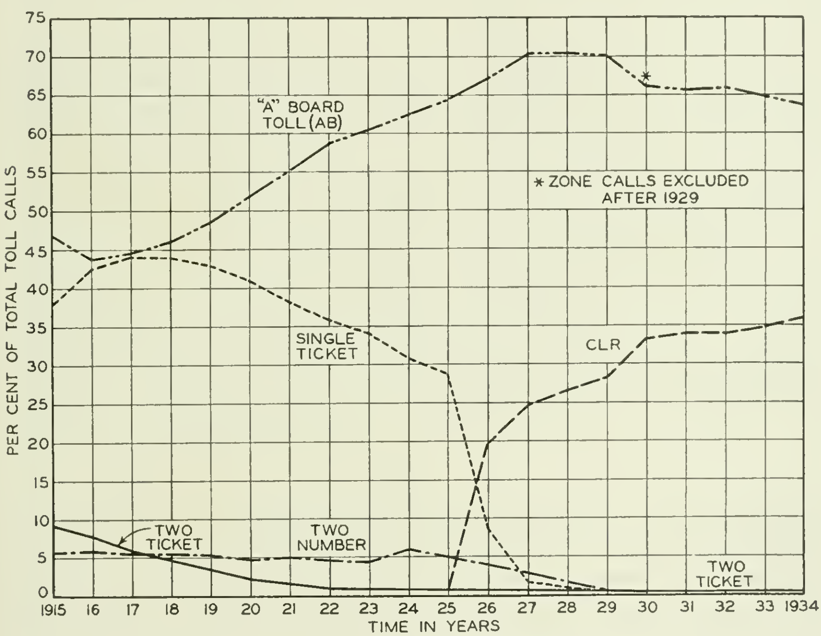


Fig. 20—Distribution of toll calls by operating methods at Bell operated offices in the United States.

present about 99 per cent are handled by the A-Board Toll and Combined Line and Recording methods. A large percentage of toll calls involve short distances and are of a simple type that may readily be handled at local switchboards. The remaining toll calls are handled

at separate long distance offices where equipment especially designed for the handling of long distance calls is provided. In addition to the switchboard positions at which the operators work who handle the long distance calls, there are certain auxiliary positions where operators supply information to other operators with regard to routes, rates and charges. Whether the toll call is handled at the local office or at the long distance office the operator who handles the call and deals with the customers must have access to the toll lines and means of communication with the auxiliary operators. Under any toll operating method the operator who handles the call must make a record from which the customer is billed. This is done on a small ticket which also serves other supervisory purposes. There must be timing devices by which the operator may time the length of conversation and facilities for sending the ticket to file or to other operators if additional or special work is to be done in connection with it.

A-Board Toll Operating Method

On the toll calls handled at local A boards, the subscriber reaches his local operator (by dialing the code "0" in dial areas) and gives his call to her. If she does not know the route to the called place from memory, she obtains it either from a bulletin at her position or by inquiry of the route operator. She reaches the called place by whatever trunking method is in use over the route in question. Most such calls today are completed by the straightforward or dialing methods over direct trunks. If direct circuits to the called place are not provided and the straightforward method is used, the operator selects a trunk to an intermediate or tandem office and upon receipt of proper signal passes the proper order to the tandem or intermediate operator who connects the trunk on which the order was received to a trunk to the called office. Upon receipt of the order at the called office, the terminating operator makes connection to the called line. If the dialing method is involved, the originating operator dials the called number over tandem trunks to the called place. When the called telephone answers, the operator enters the connect time on the ticket. When both parties hang up the receivers this action operates signals before the operator which indicate to her that the customers have finished talking. When these signals appear, the operator enters the disconnect time on the ticket above the time previously shown and the duration of the connection then may be obtained simply by subtracting the connect time from the disconnect time. If the calling party requests the charge for the call, the operator makes this subtraction, obtains the rate to the called place either from her bulletin or from the rate operator, computes the charge and advises the customer.

Two-Number Toll Operating Method

For transmission reasons it was found desirable in large metropolitan areas to provide a separate trunk plant and a separate toll switchboard for handling calls between widely separated central offices. This separate office became known as the "two-number" office and the traffic was handled by the "two-number" method. This method was used for a number of years before the present tandem systems came into service.

Under the two-number operating method the subscriber gave the called number to his local operator as on a local call. The local operator, over a trunk to the two-number board, passed the called number and then the calling number to the two-number operator. The two-number operator then obtained connection to the calling number over another trunk which afforded better transmission than that of the first trunk, and disconnected from the trunk over which the call was received. This disconnection caused a signal to light before the local operator who originally received the call, whereupon the local operator took down the connection she had made between the calling party and the two-number operator. The two-number operator then proceeded to establish connection with the called telephone over a trunk of proper transmission design by whatever trunking method was in use. At the time the two-number toll operating method was in greatest use, the usual trunking method was call-circuit although much of this business was handled by the ringdown method. Tickets were written, connections timed, and routes and rates were obtained in much the same manner that they are obtained with the A board toll operating method. The two-number method acquired its name through the fact that the local operator passed two numbers on each call to the so-called two-number operator.

Two-Ticket Toll Operating Method

For that portion of the toll business on which the customer reaches and gives his call to the long distance operator, three important toll operating methods have been used. One of the important early toll operating methods involved the writing of a ticket by the operators at both ends of the connection and became known as the "two-ticket" method.

In each long distance office where the two-ticket operating method was in use, there was provided a recording board at which long distance calls were recorded by a special group of recording operators; there was an arrangement for sending the tickets either by mechanical device or by messenger from the recording board to other positions as

required. There was a directory desk where the directory operator wrote on the ticket the telephone number of the called person. Usually associated with the directory desk was an arrangement for filing completed tickets such that should any customer wish to inquire the charge on his call after it had been filed, it might be located quickly. There was a route and rate desk to which the ticket was then sent and where the operator recorded upon it the route and the rate to the called place. There was an outward or line board where operators established connection between the calling and called telephones. There was a special board known as an inward board where operators established connections to local offices for operators at distant offices. There was a through board where operators connected toll circuits together, end to end, on calls coming from a distant city and going to another city via this office. Each pair of line positions was equipped with a device for timing calls, the calculagraph.

With two-ticket operation, a customer wishing to place a long distance call reached his local operator and asked her to connect him with long distance. The local operator complied with this request by making connection to a trunk to long distance which appeared for answering before a special group of operators trained only to record the customer's order. The recording operator answered the signal on this trunk by saying "Long Distance." The customer told the recording operator whom he wished to reach and where he might be found. He was then told by the recorder that the operator would call him and he hung up his receiver. After recording the information supplied by the customer on an "outward" ticket form, the ticket was sent to other operators for further handling. If the customer had not supplied the number of the called telephone, the ticket was sent to a directory operator who looked up in the directory of the called place the telephone number of the called person. Each toll office did not then, nor does it now, have direct circuits to all other toll offices. It was necessary, therefore, in many cases, to determine the route to the called place. After the telephone number had been supplied to the ticket by the directory operator, the ticket next went to the routing operator who indicated on the ticket the route to the called place. The ticket was then sent to the particular line operator who handled calls to the desired place. The line operator obtained connection with the calling subscriber's telephone and to a circuit to, or in the direction of, the called place. Having reached the inward operator at the called place, she passed the details of the call to her. The inward operator recorded them on an "inward" ticket form and proceeded to obtain connection with the called telephone or party or to find out

where or when the called person might be reached. Having reached the called station or party she notified the originating operator who then rang the calling party. When both the calling and called parties answered their telephones, the operator inserted the ticket in the calculagraph and stamped the time. When the calling party hung up his receiver, the originating operator received a disconnect signal on his line, stamped the time on the ticket by means of the calculagraph and took down the connection. The ticket was then sent to the ticket filing desk where it was filed in the numerical order of the calling number. The inward operator also took down the connection upon receipt of a signal indicating that the called party had hung up his receiver.

Single-Ticket Operating Method

Before the single-ticket method came into use operating methods and practices, as well as accounting methods, varied from place to place. This made it necessary for a considerable part of the operating work on toll calls to be done by the operator at the called place. The first step in passing from the two-ticket to the single-ticket method was to eliminate the ticket at the inward end, and to place the responsibility for all work in connection with handling the call, except the purely mechanical operation of making physical connection to the called telephone, upon the outward operator at the calling place. The elimination of this work made possible also the elimination of a large amount of equipment at the terminating place, avoided the duplication of operator time at both ends of the circuit and saved circuit time. It required standardization throughout the System in equipment, local and toll practices, and auditing methods.

Under the single-ticket method the customer reached long distance just as he did with the two-ticket method and the preliminary work of finding the called number and the route for the call remained unchanged. When the ticket reached the line operator, however, she took up a circuit to the called place and merely passed an order to the inward operator for connection to the called number. When the called telephone answered, the originating operator announced the call, arranged for the called party to come to the telephone and connected the calling party to the circuit when the person at the called station was ready to talk. The connection was timed and the ticket filed in the same way as under the two-ticket method.

Combined Line and Recording (CLR) Method

Experience with the single-ticket operating method had suggested the possibility of having the line operator receive and record the call

as well as perform the work of reaching the called telephone or party and of establishing the connection. Improvements in toll plant contributed to the feasibility of this type of operation. Such a plan would eliminate the need for a separate recording board but would increase the number of outward line positions required. It also would bring in new problems of training and supervision. With the proposed method it appeared that the speed of service on long distance calls might be considerably improved by virtue of the fact that it would no longer be necessary to send tickets from one position to another within the office. Furthermore, with this method it would be unnecessary for the operator to dismiss the customer after he had given her his call and to recall him when ready with the connection. Instead the customer could remain at the telephone while the line operator attempted to complete his call.

Under the CLR method of operating, now in use, the customer dials or asks for long distance in the usual way. The signal at the long distance board appears before the line operator who answers with the words "Long Distance." The line operator records the call in the usual way except that when the customer gives the name of the called place and the number of the called telephone she takes up a circuit to the called place and records the information on the ticket while waiting for the inward operator at the called place to answer. After passing the called number to the inward operator and while waiting for the called telephone to answer, she asks the calling party for his telephone number. Conversation is timed and the ticket disposed of in the usual way.

Under the single-ticket method calls to or via a given city are always handled by the same group of operators. Under the CLR method any line operator may handle a call to any place in the toll system. If the call is not completed on the first attempt, the ticket is sent to the so-called point-to-point positions where calls to a given city are assigned to positions designated to handle calls only to that city. This assures prompt and careful handling of those calls which have encountered delay and the handling of such calls does not interfere with the handling of work on new calls at the CLR position.

It may be of interest to follow the handling of a call by the CLR method. Figure 21 shows schematically the route of a long distance call through the telephone plant, and the functions performed by the various operators along the route while handling the call by the CLR method. Let us assume that the call in question is a station-to-station paid call from an individual line dial telephone in New York to an individual line telephone in Chicago. The New York subscriber

removes the receiver from its hook, listens for dial tone and dials a code for the long distance operator (in New York "211"). A signal appears before the line operator at the long distance switchboard and the customer hears the ringing signal. The line operator plugs into the trunk on which the customer's signal has appeared and indicates her readiness to receive the call by saying "Long Distance." The customer gives his order by saying "Chicago, Harrison 1234." The long distance operator plugs into a Chicago circuit with the other end of the cord pair used in answering the subscriber and rings. While

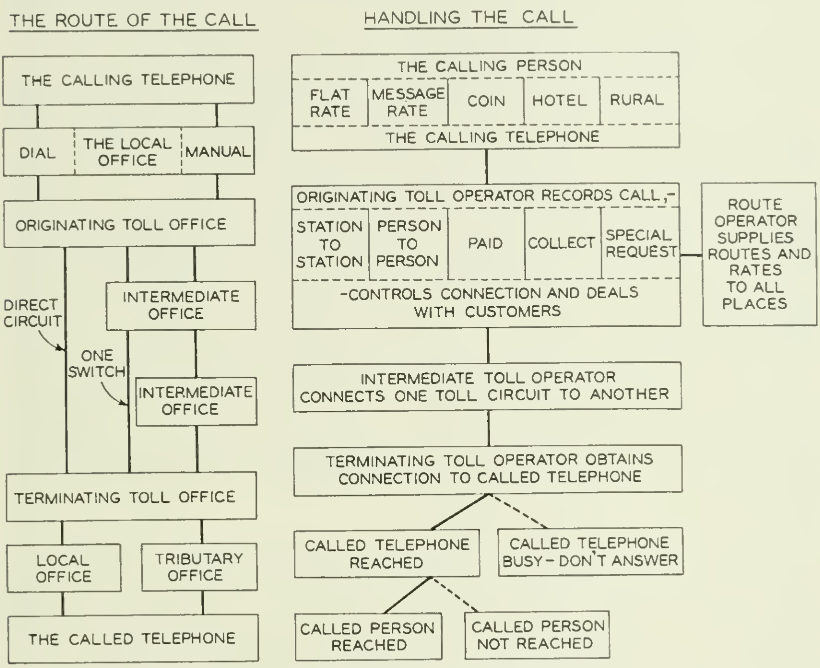


Fig. 21—The long distance call.

the customer was speaking and while performing these operations and waiting for the Chicago inward operator to answer, she has recorded the abbreviation for Chicago and the Chicago telephone number on a toll ticket. The Chicago inward operator answers by saying "Chicago." The New York operator responds with "Harrison 1234" and while waiting for the Chicago inward operator to obtain connection to this number, she asks the New York subscriber for his telephone number. The subscriber at Harrison 1234 answers and conversation begins. After recording the New York number on the toll ticket, the

New York operator inserts the ticket in the calculagraph in readiness to stamp the start of conversation. When she hears the party at Chicago speak to the New York party, she cuts out of the connection, stamps the start of conversation on the ticket and places the ticket in a clip associated with the pair of cords on which the conversation is taking place. When the parties have finished speaking, the New York party hangs up his receiver which lights a signal associated with this pair of cords, whereupon the operator inserts the ticket in the calculagraph and stamps the finish of conversation. She then takes down the connection and sends the ticket to file. When the Chicago party hangs up his receiver, the inward operator at Chicago receives a disconnect signal and likewise takes down the connection.

The above describes the steps in the handling of the simplest type of long distance call. There are many variations from this. The process varies with the type of telephone at which the call originates and at which it terminates. Person-to-person calls involve reaching particular persons and introduce additional variations in handling. Calls may be placed either paid or collect and the routine is different in each case. Direct circuits are not provided to all places and intermediate operators are involved in handling switched calls. The called telephone sometimes is busy or does not answer or the called person may not be available and additional attempts must be made to complete the call. All of these conditions call for variations in the process of handling the call, yet the operating method and the operating rules or practices which describe it must cover all of these situations.

Evolution of Toll Operating Methods

It may be seen from the above that toll operating methods have grown and developed along with the business to meet the changing requirements. As each new method came into use the quality and usefulness of the toll service has steadily improved and the way has been cleared for a better operating job and improved supervision. High grade operating and supervision broaden the possibilities in methods betterment work and may well be the controlling factors in the success of improved plans. The operating method is influenced by features of plant design such as transmission requirements and improvements in switchboard equipment. Conversely, the design and arrangement of plant is influenced by changes in operating method made to improve the quality of the service. Through the toll system the telephone service of the country as a whole is tied together as one great network and the coordination and standardization of the plant and methods make possible universal service.

General Toll Switching Plan

The technical developments which are outlined in the preceding sections of this account made possible continued improvement in the range and quality of telephone conversations over long distances and economies in the costs of providing long distance circuits. As a result, long distance telephone service grew rapidly, both in volume and in extent, and by 1915 service was established between the Atlantic and the Pacific Coasts. The application of technical developments continued, increasing the transmission efficiency not only of the very long telephone circuits which technical developments have recently made possible but also of shorter toll circuits of all lengths.

Although the opening of the transcontinental line showed the possibility of establishing direct telephone service between any two points in the country, a great deal more had to be done in order to closely realize the Bell System ideal of universal service, that is, good service between any two points in the country. While a large proportion of the toll board messages (at present 80 per cent) is handled by direct circuits between the two terminal points, there is naturally a very large number of combinations of cities and towns in the country between which the telephone business is too light to justify direct circuits—in fact, these constitute a large percentage of all the combinations of places in the country. For these conditions, when a telephone connection is required, it must be established by switching together two or more telephone circuits. Some cases might require switching together a considerable number of telephone circuits, this sometimes involving difficulty and delay in establishing the connection. Also, while the telephone circuits may be so designed that, individually or in combinations of two, they provide very satisfactory transmission, in some of these cases requiring a number of switches, the combination of circuits might result in unsatisfactory transmission.

In order that universal service for the nation might practically be realized, it was necessary to provide an underlying plan for the routing of telephone calls between any two places such that the maximum number of switches necessary for building up the connection would be as low as practicable. This must apply to connections between any two points in an operating area, or other natural subdivision of the country, and also to the country as a whole. Furthermore, the plan should provide for a transmission design of toll circuits such that transmission conditions will be satisfactory on individual circuits when used for direct traffic between their terminals, and also for any combination of circuits which may be connected together in establishing a

connection between any two points. It is the purpose of the General Toll Switching Plan to provide a general design of the toll plant which meets these requirements and which, therefore, when fully effective, provides for satisfactory service between any two points in the continental United States. The Plan also covers that part of Canada served by the Bell Telephone Company of Canada. For most of the messages, where volume of business and other conditions justify, the telephone service is of course better than the minimum contemplated by the Plan as, for example, by the provision of direct circuits.

In addition, trends in the construction of toll circuits were such that there was a growing need for an underlying plan for routing toll circuits in such a way as to provide for the most economical plant design. Large numbers of additional toll circuits were required and the types of new telephone plant were such as to trend increasingly toward the concentration of large numbers of telephone circuits on a single route. This is illustrated best by the telephone cable, making possible the installation of many hundreds of circuits along the same route and in a smaller way by the application of carrier telephone systems to open-wire lines, doubling or trebling the number of circuits which could be carried by each such line. Satisfactory operation over connections built up by switching together several toll circuits (multi-switch connections as they are called) requires the insertion of transmission gain at the switching points. Developments in methods of providing such transmission gain by the proper manipulation of repeaters were of such nature that increasing economies could be realized by concentrating through switching as far as possible at a small number of points. Also, this concentration could result in operating economies. A General Toll Switching Plan lends itself naturally to concentrations of circuits on important routes, and in the development of the plan, account was taken of this trend. It therefore forms a background for realizing in future plant extensions the maximum economies from these concentrations of route and of through switching.

These considerations led to the development in 1928 and 1929 of a General Toll Switching Plan. The general features of this Plan may be understood by referring to Figs. 22 and 23. Figure 22 shows how the Plan applies within a given operating area such as an operating unit of an Associate Company. Within such an area there were selected a few important switching points and these were designated as "primary outlets." Each toll center in the area is directly connected to at least one primary outlet and each primary outlet is directly connected to every other primary outlet in the area. There-

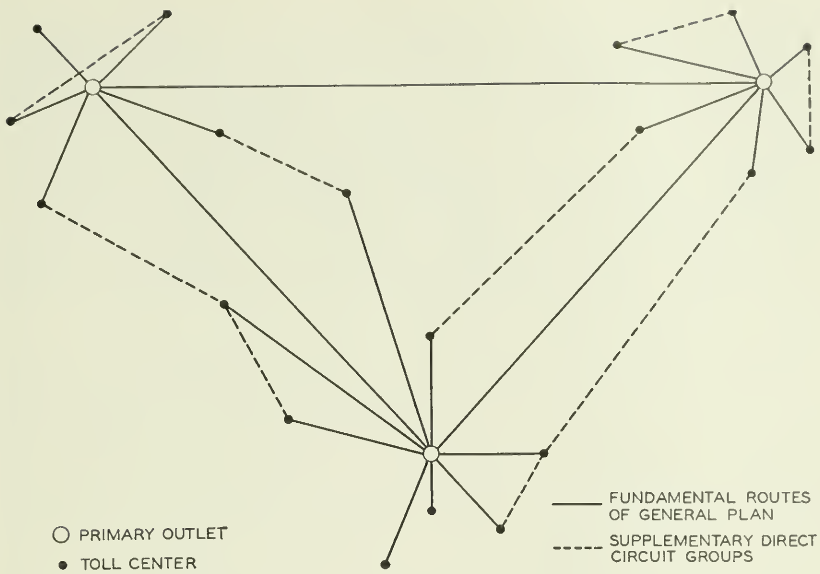


Fig. 22—Application of the toll switching plan to an operating area.

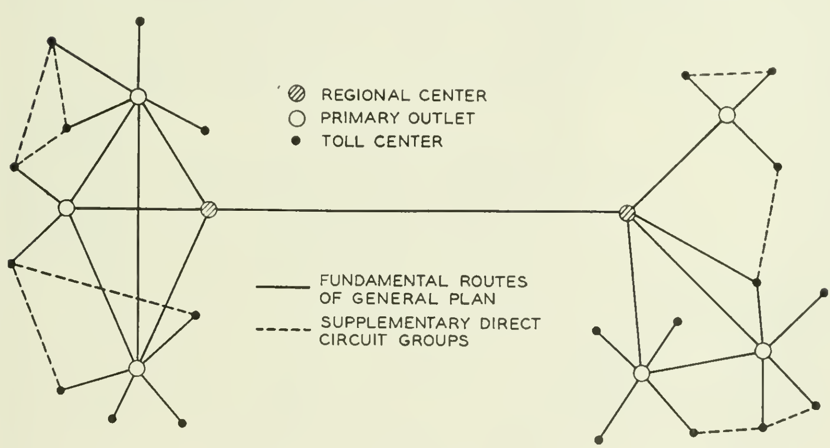


Fig. 23—Application of the toll switching plan to the country as a whole.

fore, any two toll centers in the area can be connected together with a maximum of two intermediate switches. The primary outlets for each area were selected after a careful study of present switching and operating conditions. Due weight also was given to the probable future trends. The number and location of primary outlets selected

for each operating area were designed to give maximum economy considering both present and future conditions. This naturally resulted in many cases in the selection of the larger cities of a given operating area although in some cases other points were chosen as primary outlets due to their advantageous location, for example, at the point of intersection of a number of important toll routes. The routings provided by the Plan are supplemented by direct routes or other routings where the volume of traffic or other conditions made this desirable. These other routings, however, are designed to provide service conditions at least as good as those provided by the General Toll Switching Plan.

Figure 23 shows the application of the General Toll Switching Plan to the country as a whole. In order to tie together with a minimum number of switches the interconnected groups of primary outlets, each one of these primary outlets has direct connection to at least one very important switching point designated as a "regional center," and each regional center has direct connection to every other regional center in the country. This means that any two primary outlets in the country are connected together with a maximum of two intermediate switches. The numbers of switches between telephone points of different classifications are shown by Fig. 24. It will be noted that in the limiting

To— / From	Same Regional Area				Another Regional Area			
	Regional Center	Primary Outlet	Toll Center Directly Connected to Regional Center	Toll Center Directly Connected to Primary Outlet	Regional Center	Primary Outlet	Toll Center Directly Connected to Regional Center	Toll Center Directly Connected to Primary Outlet
Regional Center	0	0	0	1	0	1	1	2
Primary Outlet	0	1	1	2	1	2	2	3
Toll Center (directly connected to Regional Center)	0	1	1	2	1	2	2	3
Toll Center (directly connected to Primary Outlet)	1	2	2	3	2	3	3	4

Fig. 24—Numbers of switches between telephone points of different classifications.

case of toll centers in different regional areas and not connected directly to any regional center, the maximum number of intermediate switches is four.

The present location of regional centers and primary outlets in the General Toll Switching Plan is shown in Fig. 25. There are, as will

Fig. 25—Locations of primary outlets and regional centers.

be noted, eight regional centers in the United States, and 143 primary outlets including three in the eastern part of Canada.

While a regional center has direct circuits to all of the primary outlets tributary to it, it also has direct circuits to many other primary outlets. This is illustrated by Fig. 26 showing for the case of the Chicago regional center the direct circuits to a large number of primary outlets throughout the country. In addition, Chicago has, of course, direct circuits to many toll centers which are not primary outlets where the volume of traffic is sufficient to justify such direct circuits. The same is true also of other regional centers and of primary outlets.

In order that the transmission between any two points in the country over a circuit routed in accordance with the General Toll Switching Plan should be satisfactory, standards were established for each class of toll circuit, that is, for toll circuits between toll centers and primary outlets, between primary outlets and regional centers, etc. These standards provide satisfactory overall transmission for connections between any two points in an operating area, with an economical division of the total transmission loss between the different toll circuits entering into the connection. Generally speaking, these same circuits form parts also of very long connections, switching at primary outlets or regional centers to long circuits running to other parts of the country. In order that satisfactory transmission may be given under these conditions, it is necessary that severe requirements be applied to the very long circuits with the result that they must be designed and maintained with great care and coordination throughout their entire length. It is also necessary that transmission gain be inserted at points where circuits are connected together, and therefore that the characteristics of the shorter circuits be such that they do not limit the possibilities of inserting such transmission gains. The application of these various complex requirements for toll circuits, in order that they may form satisfactory links in any connection, short or long, in the nation-wide toll telephone network, is greatly facilitated by the systematic character of the General Toll Switching Plan, and by the recommendations as to minimum performance standards which that plan contains.

Until recently, the method generally used for inserting transmission gain on through connections of toll circuits was by means of repeaters associated with the cord circuits at the intermediate switching points. About the time that the Toll Switching Plan was established, there was made available an improved method by which the gain of repeaters permanently inserted in the toll line is automatically adjusted at the switching point when the toll circuits are connected together. These

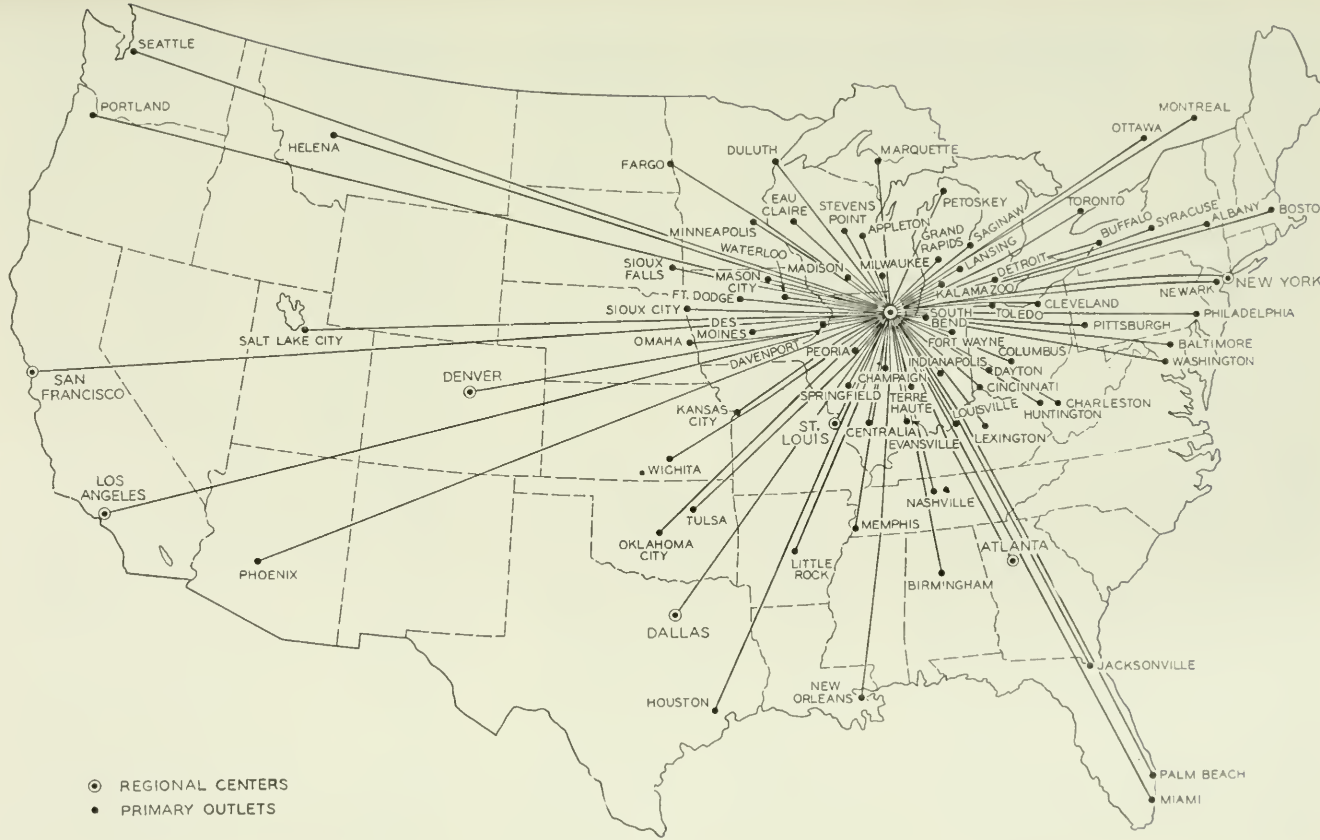


Fig. 26—Direct circuits from Chicago to primary outlets and regional centers.

improved arrangements were scheduled for application as circumstances warranted to the regional centers and primary outlets. At the present time they have been applied to all of the regional centers and about half of the primary outlets and about 95 per cent of the switched connections requiring gain at the switching centers make use entirely of this improved method.

As the Bell System is a living and growing organism, the General Toll Switching Plan is continuously under review and frequently revised in detail. For example, while there are now four primary outlets in the territory of the New England Telephone and Telegraph Company, it seems probable that future developments in concentration of circuits on cable routes will result in reducing these in number.

The transmission requirements applied to the different classifications of circuit for best results vary with the availability of further technical developments. For example, in the future the improved high-speed circuits made possible by the application of carrier to cables will result in modifications of the General Toll Switching Plan, resulting in improvements of transmission over all switched connections and in economies in circuit design through liberalizing the transmission requirements for certain routes, particularly the circuits between toll centers and primary outlets.

The General Toll Switching Plan is an important instrument in systematizing plans for the design of plant extensions, for the application of technical and operating improvements, and for realizing in fact the ideal of universal service between any two telephones in the country.

The Joint Occupancy Of Plant

The Bell System organizations involved directly in the giving of telephone service include regional companies (known as the Associate Companies), operating in various areas throughout the United States, who are responsible for the exchange service and for toll service within their areas,* and the American Telephone and Telegraph Company, responsible for the long distance toll service between points in the areas of different regional companies.

As a result, it is a common situation to have inter-area toll plant of one company terminating in towns and cities where the exchange plant is owned and operated by another company, and sometimes extending for considerable distances along the same general routes as the intra-area toll plant of that company. In a great many cases

* There are a few companies in which certain interstate items of traffic within the company area are handled by the American Telephone and Telegraph Company.

there are economic and service advantages in the consolidated construction of plant used by the two companies involved and, for such situations, this is the common practice. This involves the joint use of land, buildings, right of way, pole lines, conduits, etc.

The economic advantages of such joint use of plant are obvious. For example, one duct run with a sufficient number of ducts for both companies can be built more cheaply than two separate duct runs, and the same is true of pole lines. When toll cable is installed, it is an economy, when practicable, to place within one sheath sufficient circuits to take care of the requirements of both companies. The same is true for other parts of the telephone plant. Furthermore, there are advantages in having the toll switchboard in a building used for exchange service and often in having inter-area and intra-area toll circuits terminate at the same switchboards and use the same groups of trunk to the local exchange plant.

Other things being equal, there is an advantage in each company owning the plant required for its service, and this is the basis generally followed. This leads to a large extent to the joint ownership of jointly used plant, particularly of outside plant. This includes joint conduit runs, joint pole lines, and jointly owned toll cables.

In some cases, rather than joint ownership of jointly used plant, there are advantages in a single ownership by one of the companies, generally the company having the largest requirements, which leases some of the plant to the other company. This applies, for example, to land and buildings where, because of the greater ease and simplicity of transactions of various sorts, a single ownership is preferable. This is also often true where the requirements of one company are small or where, as a result of growth, the division of ownership of jointly occupied plant no longer corresponds exactly to the relative needs of the two companies for the use of the plant. It applies also to the temporary use by one company of spare plant owned by another which will later be required for the owner's use.

Rental Arrangements

To meet the varying conditions, two general bases of rental are in use: (1) the "reserved plant" basis, and (2) the "spare plant" basis.

On the reserved plant basis, rentals cover plant designed and constructed by the owning company for joint use with the renting company or for the sole use of the renting company under a specific plan mutually agreed upon, usually in advance of construction. Existing plant may also be put on a reserved basis by specific agreement. The plant reserved for the lessee provides not only for its

present needs, but usually for growth as well. A good example of this type of arrangement is a jointly occupied building which is designed to meet the present and expected future requirements of both companies and in which certain designated space is reserved for the lessee company.

The spare plant basis applies to plant which, in general, the owning company has provided for its own use in anticipation of its own future requirements, or to plant which is spare because of fluctuating load demands and can be temporarily placed at the disposal of the renting company. Such plant may be released by the lessee at any time or may be taken back by the lessor company at any time upon reasonable notice to the other company.

Illustrations of both these bases of rental where the American Telephone and Telegraph Company and an Associate Company are involved are given below.

(*a*) *Buildings*

In providing building space, it is the general practice for one company to own the building used jointly by both companies. This arrangement is advantageous, particularly in the larger cities, since it permits one company to deal with taxing authorities, zoning commissions, public works authorities, and the public generally. Where space for a local central office is required, it is the general practice for the Associate Company to own the building. The American Telephone Company's ownership in buildings is accordingly largely confined to intermediate repeater stations. The owning company generally furnishes space to the other company on a reserved plant basis.

(*b*) *Equipment*

In the case of toll equipment, one of several arrangements is followed, depending in part upon local conditions, such as local operating or maintenance conditions, and in part upon the relative amount of equipment required by the two companies involved. In some cases where one company requires a relatively large part of the total equipment used, that company owns all equipment and furnishes equipment for the other company's needs on a reserved rental basis. Thus, on some of the long through routes where the American Telephone and Telegraph Company uses the majority of the equipment in the intermediate repeater stations, it owns all equipment and rents such portion as required to meet the other company's needs on a reserved rental basis. In many other cases where both the American Telephone and Telegraph Company and the Associate Company have considerable toll equipment requirements and these can best be provided in joint installation, each company will own the equipment provided for its use. If there is any sudden peak in the equipment requirements of one company, the other company will usually temporarily furnish spare equipment from its own reservation on a rental basis to aid in meeting the peak demands.

(c) *Outside Plant*

Where both companies' use is substantial, arrangements are usually made for the joint ownership of the common items of plant. In underground conduit, the ownership is usually divided on the basis of number of ducts required by each company, except that if one company's requirements are less than one-half of one duct (occupied by a jointly owned cable), it is customary for that company to rent duct space from the other company. Open-wire pole lines are generally jointly owned where each company has a requirement of one crossarm or more, the cost of the pole line being divided in proportion to the numbers of crossarms required by each company. In cases where the requirements of one company are minor, it may lease space from the other company on an attachment rental basis, using a reciprocal rental rate for the use of the supporting structure which reflects the average carrying charges on both line and right of way. In the case of toll cables, the ownership of certain wires in the cable is generally held by each company, the cost of the cable being divided in proportion to the copper cross-section of the wires owned by each of the companies.

In the open-wire plant, each pair or phantom group is generally owned by one company and located in the crossarm space reserved on the pole line for that company.

Emergency situations arise from time to time in which service may be restored most quickly by a temporary use of spare facilities of the other company or by a temporary pooling of the circuits of both companies which remain in service and applying them most equitably to the service demands of both companies. The work of restoring service in such cases is handled without the execution of any formal agreements between the companies involved, and such adjustments as are necessary are worked out later.

Joint Maintenance Arrangements

The joint maintenance arrangements are based on the principle of providing the most economical procedure in each case. This results, generally speaking, in the maintenance by employees of one company of all jointly occupied outside plant on a single route. It is obviously economical, for example, to have such an arrangement for the maintenance of pole lines, conduit, and cables which are jointly owned by the two companies.

In the case of central office equipment, it is generally desirable in large cities where a large amount of equipment is owned by each company to have separate maintenance staffs, particularly for the service maintenance work performed by the toll test room forces. At smaller points, a single maintenance force is generally provided by the company having the greater amount of work.

The division of maintenance costs between the two companies is based upon the same principles of equitable allocation as applied to the division of ownership and to rental charges. For example, the cost of maintaining toll cables is divided between the companies in proportion to their ownership interest in the cable. Another example is the cost of pole replacement, which is one of the large items of pole line costs. Replacements are made upon the basis of periodic inspections of the pole line, the first inspection being made about ten years after the new line is built and subsequent inspections approximately every four years. These inspections determine the poles which are in such deteriorated condition as to require replacement. Where the replacements consist of substituting the same size of pole for the existing pole, the charges are borne by the companies concerned on the basis of their assignments on the old pole.

General

The above indicates briefly the types of arrangements for the usual case. No attempt has been made, however, to indicate all of the variations in these arrangements applying to the extensive plant of the Bell System covering the entire country and sometimes requiring modifications of these arrangements or some other special provisions. However, the general principle outlined above is followed, namely, that of providing the most economical overall result with an equitable division of costs and responsibility between the companies involved in each case.

Standardization

The electrical design of telephone toll circuits is necessarily complicated, as the overall electrical characteristics on which the efficiency of the circuits depends are the result of the composite effect of many different electrical phenomena. Also, the overall characteristics of a toll circuit are the composite resultant of the characteristics of a large number of individual pieces of apparatus and sections of circuit. The construction of the plant at such times and in such quantities as to produce most economic results involves many considerations. In many cases, as for example, in the construction of pole lines and of toll cables, it is necessary for greatest economy to provide plant to meet the estimated requirements for a considerable period ahead. Furthermore, the maintenance of this plant at a higher degree of efficiency and its operation to connect together quickly and accurately any two of the fourteen million telephones in the Bell System involve a good deal of complication in routines and procedures. In view of

these considerations, the telephone toll plant and service offer very good examples of the advantages of standardization in plant and in operating methods.

The complexity of the toll plant and the number of types of apparatus and material which would be required would be greatly multiplied if it were not for the high degree of standardization in the Bell System telephone plant. In fact, it is not an exaggeration to say that the telephone toll service of today could not be given had not effective steps been taken from the beginning looking to this high degree of standardization and simplification.

Plant Design

It is evident that if each toll circuit were designed individually to meet exactly the requirements for that circuit as regards efficiency for good transmission and other requirements, the result would be, in general, that each small group of toll circuits between two points would differ in electrical design from every other group of toll circuits between any other two points. This would result in many thousands of different kinds of toll facilities, each designed for a specific use only, and would result in endless confusion and lack of practicability. However, the standardization of the apparatus and materials forming the toll telephone plant has been carried on since the beginning of toll service and has resulted in a simplification of practice and the general use of the same types of apparatus and material throughout the country. For example, there has been a high degree of standardization of the sizes of copper wire used for open-wire telephone conductors. A very large percentage of the wire used in the plant for this purpose is made up of three sizes, respectively, 104, 128, and 168 mils in diameter. In toll cables, practically all conductors are made up entirely of two gauges, 16 and 19 B & S gauge. With few exceptions, repeaters for telephone message circuits are of either one of two basic types, one for two-wire circuits and one for four-wire circuits, with such modifications in balancing arrangements, signaling arrangements, etc., as are necessary to adapt them to the different types of circuit. Carrier systems are one of two general types, a three-channel system for long distances and a single-channel system for shorter distances, although additional types of system for other types of circuit condition are now under development. In the design of any given circuit, choice is made from this limited number of types of facilities, selecting the one which will give not less than the required transmission efficiency in the given case with maximum economy and other advantages. This procedure results in great advantages in simplicity of plant design and

in the flexibility with which sections of toll circuit can be transferred from one use to another as occasion requires.

Construction, Maintenance, and Operation

The advantages of standardization apply to the operating field as well as to the engineering design. Standard construction practices are based upon the use of standard types of construction material all over the country. This standardization of materials makes it possible for the purchasing organization to buy large quantities of a relatively small number of types of material with a resulting saving in cost. Also, the standard construction practices facilitate the training of men and the transfer of men from one part of the System to another with shifting needs.

Similar advantages result from the standardization of maintenance practices. The Bell System maintenance practices make use of the maintenance experiences of the operating companies and of general investigations of the relative advantages of different practices and methods. These practices are generally used throughout the country with advantages from the standpoints both of economy and of service. Men at widely separated points and sometimes employed by different companies can cooperate closely in the maintenance of telephone circuits which have been or may be connected together in the toll service. Also, at times of emergency, men and materials from various parts of the country, wherever available, may be concentrated on the emergency job, and the men, applying standard methods to standard materials with which they are familiar, can work most effectively in the quick restoration of service.

It is perhaps in considering traffic operation practices that the advantages and, indeed, the necessity of standardization in relation to operation is most evident. The toll operators must constantly deal with other operators in distant cities, and it is obviously essential that the operating practices should be alike in order to avoid extreme difficulties and reaction on the speed and quality of service. The standard Bell System operating practices provide in detail the standard procedures to be followed by operators in handling the various types of toll call and, in general, specify also the phraseology to be used by operators with a view to insuring maximum accuracy, clearness, and convenience to subscribers.

General

While, as pointed out in the above paragraphs, standardization in the Bell System is a means of obtaining economy and efficiency, it is more than that. It is essential to the best service and the most rapid

progress. The conditions bearing on the telephone toll plant and toll service are constantly changing through growth, shifting demands, the development of new needs of customers for telephone service. Also, the best means and methods available for giving service are constantly developing as a result of the experience of the various operating companies and through the development of new instrumentalities and operating methods by the headquarters forces of the Bell System. These types of apparatus and of communication systems, and methods and practices for construction, maintenance or operation represent the outcome of careful consideration of the best way to meet a type of situation. Through the headquarters organization their availability is made known at once to the operating telephone companies of the Bell System throughout the country with information regarding their desirable field of use. This greatly facilitates their adoption and application by these Companies.

In some cases, such new standards present means for doing something which could not be done before. In many cases, such new standards replace existing standards due to advances in the art, improvements in methods or technique, or changes in operating requirements. Standardization in the Bell System, therefore, involves a continuous procession of new standards to meet new conditions or to meet old conditions better than was heretofore possible, and the subsequent dropping of old standards. Such standardization is based not only upon the present needs of the telephone system, but also upon the best picture which can be formed of future trends. It is essential to the rapid and satisfactory development of telephone toll service.

Bibliography

Page Number of Statement	*Title*	*Author*
	Early Developments:	
2–15	"Beginnings of Telephony" N. Y., Harper and Brothers, 1929, 261 pp.	F. L. Rhodes
2–15	"Some Recent Advances in Telephony" *A. I. E. E. Transactions*, Vol. III, 1886, p. 74–93.	Thomas D. Lockwood
2–3	"Speech and Hearing" N. Y., D. Van Nostrand Co., Inc., 1929, 331 pp.	Harvey Fletcher
2	"The Development of the Microphone" *Bell Telephone Quarterly*, Vol. 10, July 1931, pp. 164–188.	H. A. Frederick
3	"The Carbon Microphone" *Journal of the Franklin Institute*, Vol. 217, April 1934, pp. 407–442; *Bell System Technical Journal*, Vol. 13, April 1934, pp. 163–194.	F. S. Goucher

Page Number of Statement	*Title*	*Author*
4–5	"Telephone Switchboard—Fifty Years of History" *Bell Telephone Quarterly*, Vol. 7, July 1928, pp. 149–165.	F. B. Jewett
5–6	"The Modern Telephone Cable" *A. I. E. E. Transactions*, Vol. XXVIII, July 1909, pp. 1079–1093.	F. B. Jewett
5–6	"Development of Cables Used in the Bell System" *Bell Telephone Quarterly*, Vol. 2, April 1923, pp. 94–106.	F. L. Rhodes
6–8	"Inductive Disturbances in Telephone Circuits" *A. I. E. E. Transactions*, Vol. VIII, March 1891, pp. 100–108.	J. J. Carty
13–14	"Propagation of Long Electrical Waves" *A. I. E. E. Transactions*, Vol. XVI, March 22, 1899, pp. 93–142.	M. I. Pupin
13–14	"Wave Transmission over Non-uniform Cables and Long Distance Air-Lines" *A. I. E. E. Transactions*, Vol. XVII, May 19, 1900, pp. 445–507.	M. I. Pupin
13–14	"On Loaded Lines in Telephonic Transmission" *Philosophical Magazine*, Vol. 5, Mar. 1903, pp. 313–330.	G. A. Campbell
13–15	"Development and Application of Loading for Telephone Circuits" *Bell System Technical Journal*, Vol. 5, April 1926, pp. 221–281; *A. I. E. E. Transactions*, Vol. 45, Feb. 1926, pp. 268–292; *Electrical Communication*, Vol. 4, April 1926, pp. 258–276; July 1926, pp. 38–53.	Thomas Shaw William Fondiller
13–15	"Permalloy; The Latest Step in the Evolution of the Loading Coil" *Bell Telephone Quarterly*, Vol. 6, Oct. 1927, pp. 239–246.	F. L. Rhodes
13–15	"New York–Chicago Telephone Circuits: Forty Years of Growth and Progress" *Bell Telephone Quarterly*, Vol. XI, Oct. 1932, pp. 301–308.	A. B. Clark
13–15	"Propagation of Periodic Currents Over a System of Parallel Wires" *Bell System Technical Journal*, Vol. VI, pp 495–545, July 1927.	J. R. Carson R. S. Hoyt
14–15	"Commercial Loading of Telephone Circuits in the Bell System" *A. I. E. E. Transactions*, Vol. 30, pt. 3, June 1911, pp. 1743–1764.	Bancroft Gherardi

Page Number of Statement	*Title*	*Author*
	Possibilities and Problems Associated with the Use of the Telephone Repeater:	
15–22	"Telephone Repeaters" *A. I. E. E. Transactions*, Vol. 38, pt. 2, Oct. 1919, pp. 1287–1345.	Bancroft Gherardi F. B. Jewett
15–22	"Telephone Repeaters" *Electrical Communication*, Vol. 1, Aug. 1922, pp. 6–10; Nov. 1922, pp. 27–36.	Bancroft Gherardi
16–17	"The Audion. A New Receiver for Wireless Telegraphy" *A. I. E. E. Transactions*, Vol. XXV, Oct. 26, 1906, pp. 735–763.	Lee De Forest
16–17	"Thermionic Vacuum Tubes and Their Applications" *Bell System Technical Journal*, Vol. 2, Oct. 1923, pp. 31–100.	R. W. King
16–17	"The Vacuum Tube in Telephone Communication" *Bell Telephone Quarterly*, Vol. 12, Oct. 1933, pp. 240–250.	D. K. Gannett
16	"Practical Application of the Telephone Repeater" *The Western Society of Engineers Journal*, Vol. 27, May 1922, pp. 129–142.	H. S. Osborne
18–20	"Limitation of the Gain of Two-Way Telephone Repeaters by Impedance Irregularities" *Bell System Technical Journal*, Vol. 4, Jan. 1925, pp. 15–25.	George Crisson
18–20	"Irregularities in Loaded Telephone Circuits" *Bell System Technical Journal*, Vol. 4, Oct. 1925, pp. 561–585.	George Crisson
21–22	"The Long Struggle Against Cable Crosstalk" *Bell Telephone Quarterly*, Vol. XIV, Jan. 1935, pp. 3–12.	M. A. Weaver
22	"Open-Wire Crosstalk" *Bell System Technical Journal*, Vol. XIII, pp. 19–58, Jan. 1934, pp. 195–238, Apr. 1934.	A. G. Chapman
22	"Certain Factors Limiting the Volume Efficiency of Repeatered Telephone Circuits" *Bell System Technical Journal*, Vol. XII, pp. 517–532, Oct. 1935.	L. G. Abraham

Page Number of Statement	*Title*	*Author*
	Toll Cable System:	
22–26	"Telephone Transmission Over Long Cable Circuits" *A. I. E. E. Transactions*, Vol. 42, 1923, pp. 86–97; *Electrical Communication*, Vol. 1, Feb. 1923, pp. 26–40; *Bell System Technical Journal*, Vol. 2, Jan. 1923, pp. 67–94.	A. B. Clark
22–26	"Some Recent Developments in Long Distance Cables in the United States of America" *Bell System Technical Journal*, Vol. 9, July 1930, pp. 487–492.	A. B. Clark
22–26	"Long Distance Telephone Circuits in Cable" *Bell System Technical Journal*, Vol. 11, Oct. 1932, pp. 520–545.	A. B. Clark H. S. Osborne
23–26	"Systems of Long Telephone Lines" *Bell Telephone Quarterly*, Vol. 13, Jan. 1934, pp. 23–35.	J. J. Pilliod
25	"Long Toll Cable Construction and Maintenance" *Bell Telephone Quarterly*, Vol. VIII, No. 2, April 1929.	L. N. Stoskopf
	The Time Factor in Telephone Transmission:	
26–29	"The Time Factor in Telephone Transmission" *Bell System Technical Journal*, Vol. 11, Jan. 1932, pp. 53–66; *A. I. E. E. Transactions*, Vol. 51, Mar. 1932, pp. 141–147.	O. B. Blackwell
28	"Echo Suppressors for Long Telephone Circuits" *A. I. E. E. Transactions*, Vol. 44, 1925, pp. 481–490; *Electrical Communications*, Vol. 4, July 1925, pp. 40–50.	A. B. Clark R. C. Mathes
29	"Building-Up of Sinusoidal Currents in Long Periodically Loaded Lines" *Bell System Technical Journal*, Vol. 3, Oct. 1924, pp. 558–566.	J. R. Carson
29	"Phase Distortion in Telephone Apparatus" *Bell System Technical Journal*, Vol. 9, July 1930, pp. 493, 521.	C. E. Lane
29	"Measurement of Phase Distortion" *Bell System Technical Journal*, Vol. 9, July 1930, pp. 522–549.	Harry Nyquist Smart Brand
29	"Effects of Phase Distortion on Telephone Quality" *Bell System Technical Journal*, Vol. 9, July 1930, pp. 550–566.	J. C. Steinberg

Page Number of Statement	Title	Author
29	"Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks" *Bell System Technical Journal*, Vol. 7, July 1928, pp. 438–534.	O. J. Zobel
	Multi-Channel Telephone Systems:	
30–34	"Carrier Current Telephony and Telegraphy" *A. I. E. E. Transactions*, Vol. 40, 1921, pp. 205–300; *Electrician*, Vol. 86, Apr. 15, 1921, pp. 451–453; Apr. 22, 1921, pp. 485–487; May 6, 1921, pp. 551–554 (Abstract).	E. H. Colpitts O. B. Blackwell
30–34	"Making the Most of the Line" *Electrical Communication*, Vol. 3, July 1924, pp. 8–21.	F. B. Jewett
31–32	"Transmission Characteristics of Electric Wave Filters" *Bell System Technical Journal*, Vol. 3, Oct. 1924, pp. 567–620.	O. J. Zobel
31–33	"Carrier System on Long Distance Telephone Lines" *A. I. E. E. Transactions*, Vol. 47, Oct. 1928, pp. 1360–1386; *Bell System Technical Journal*, Vol. 7, July 1928, pp. 564–629.	H. A. Affel C. S. Demarest C. W. Green
33	"Carrier Telephone System for Short Toll Circuits" *A. I. E. E. Transactions*, Vol. 48, Jan. 1929, pp. 117–140.	H. S. Black M. L. Almquist L. M. Ilgenfritz
31–33	"Practical Application of Carrier Telephone and Telegraph in the Bell System" *Bell System Technical Journal*, Vol. 2, April 1923, pp. 41–52.	A. F. Rose
34	"Stabilized Feedback Amplifiers" *Bell System Technical Journal*, Vol. XIII, Jan. 1934, pp. 1–18.	H. S. Black
34	"Electrical Wave Filters Employing Quartz Crystals as Elements" *Bell System Technical Journal*, Vol. 13, July 1934, pp. 405–452.	W. P. Mason
34	"Communication by Carrier in Cable" *Electrical Engineering*, Vol. 52, July 1933, pp. 477–481.	A. B. Clark
	"Carrier in Cable" *Bell System Technical Journal*, Vol. 12, July 1933, pp. 251–263.	B. W. Kendall

Page Number of Statement	*Title*	*Author*
34	"Systems for Wide-Band Transmission over Coaxial Lines" *Bell System Technical Journal*, Vol. 13, Oct. 1934, pp. 654–679.	L. Espenschied
34	"Wide Band Transmission over Coaxial Lines" *Electrical Engineering*, Vol. 53, Oct. 1934, pp. 1371–1380; *Electrical Communication*, Vol. 13, Oct. 1934, pp. 159–171.	M. E. Strieby
34	"Wide-Band Transmission over Balanced Circuits" *Bell System Technical Journal*, Vol. XIV, pp. 1–7, Jan. 1935.	A. B. Clark
34	"Wide-Band Transmission in Sheathed Conductors" *Bell Telephone Quarterly*, Vol. XIV, p. 145, July 1935.	O. B. Blackwell
	Associated Technical Developments:	
34–40	"Telephone Equipment for Long Cable Circuits" *A. I. E. E. Transactions*, Vol. 42, 1923, pp. 742–752; *Bell System Technical Journal*, Vol. 2, July 1923, pp. 112–140.	C. S. Demarest
38	"Metallic Polar-Duplex Telegraph System for Long Small-Gage Cables" *A. I. E. E. Transactions*, Vol. 44, 1925, pp. 316–325; *Electrical Communication*, Vol. 3, April 1925, 276–287.	J. H. Bell R. B. Shanck D. E. Branson
38–39	"Voice-Frequency Carrier Telegraph System for Cables" *Electrical Communication*, Vol. 3, April, 1925, pp. 288–294; *A. I. E. E. Transactions*, Vol. 44, 1925, pp. 327–332.	B. P. Hamilton Harry Nyquist M. B. Long W. A. Phelps
	Development of Methods of Measurement and of Maintenance for Toll Circuits:	
40	"A Shielded Bridge for Inductive Impedance Measurements at Speech and Carrier Frequencies" *Bell System Technical Journal*, Vol. 6, Jan. 1927, pp. 142–171; *A. I. E. E. Transactions*, Vol. 45, 1926, pp. 1266–1277.	W. J. Shackelton
40	"The Location of Opens in Toll Telephone Cables" *Bell System Technical Journal*, Vol. 6, Jan. 1927, pp. 27–54.	P. G. Edwards H. W. Herrington

Page Number of Statement	Title	Author
40–43	"Measuring Methods for Maintaining the Transmission Efficiency of Telephone Circuits" *A. I. E. E. Transactions*, Vol. 43, 1924, pp. 423–433.	F. H. Best
40–43	"Electrical Tests and Their Applications in the Maintenance of Telephone Transmission" *Bell System Technical Journal*, Vol. 3, July 1924, pp. 353–392.	W. H. Harden
40–43	"Practices in Telephone Transmission Maintenance Work" *Bell System Technical Journal*, Vol. 4, Jan. 1925, pp. 26–51; *A. I. E. E. Transactions*, Vol. 43, 1924, pp. 1320–1330.	W. H. Harden
	Special Services:	
43–45	"Telephone Circuits Used as an Adjunct to Radio Broadcasting" *Electrical Communication*, Vol. 3, Jan. 1925, pp. 194–202.	H. S. Foland A. F. Rose
43–45	"Telephoning Radio Programs to the Nation" *Bell Telephone Quarterly*, Vol. 7, Jan. 1928, pp. 5–16.	L. N. Stoskopf
43–45	"Wire Line Systems for National Broadcasting" *Bell System Technical Journal*, Vol. 9, Jan. 1930, pp. 141–149; *I. R. E. Proceedings*, Vol. 17, Nov. 1929, pp. 1998–2005.	A. B. Clark
43–45	"Long Distance Cable Circuit for Program Transmission" *Bell System Technical Journal*, Vol. 9, July 1930, pp. 567–594; *A. I. E. E. Transactions*, Vol. 49, Oct. 1930, pp. 1514–1523.	A. B. Clark C. W. Green
43–45	"Wide-Band Open-Wire Program System" *Bell System Technical Journal*, Vol. 13, July 1934, pp. 361–381; *Electrical Engineering*, Vol. 53, April, 1934, pp. 550–562.	H. S. Hamilton
46	"Transmission of Pictures over Telephone Lines" *Bell System Technical Journal*, Vol. 4, April 1925, pp. 187–214.	H. E. Ives J. W. Horton R. D. Parker A. B. Clark
	Toll Operating Methods:	
49	"Operating Features of the Straightforward Trunking Method" *Bell Telephone Quarterly*, Vol. IX, No. 1, January, 1930.	F. S. James W. E. Farnham

Page Number of Statement	*Title*	*Author*
54–57	"Improvement of Toll Service Through Combined Line and Recording Method" *Bell Telephone Quarterly*, Vol. X, No. 2, April 1926.	K. W. Waterson
	General Toll Switching Plan:	
58–65	"A General Switching Plan for Telephone Toll Service" *Bell System Technical Journal*, Vol. 9, July 1930, pp. 429–447.	H. S. Osborne
	Standardization:	
69–72	"Standardization in the Bell System" *Bell Telephone Quarterly*, Vol. 8, Jan. 1929, pp. 9–24; April 1929, pp. 132–152.	H. S. Osborne
	Other Papers:	
	"Building for Service" *Bell Telephone Quarterly*, Vol. VII, April 1928, pp. 69–81.	H. P. Charlesworth
	"Telephone Communication System of the United States" *Bell System Technical Journal*, Vol. IX, Jan. 1930, pp. 1–100.	Bancroft Gherardi F. B. Jewett
	"Engineering Reports of the Joint Subcommittee on Development and Research" Volume 1, April 1930; Volume 2, April 1932.	National Electric Light Association and Bell Telephone System
	"A Survey of Room Noise in Telephone Locations" (Abridgment) *A. I. E. E. Journal*, Sept. 1930, pp. 791–795.	W. J. Williams R. G. McCurdy
	"Effects of Rectifiers on System Wave Shape" *Electrical Engineering*, Jan. 1934, pp. 54–63.	P. W. Blye H. E. Kent
	"Mutual Impedance of Ground Return Circuits. Some experimental studies" *A. I. E. E. Journal*, Aug. 1930, pp. 657–661, (Abridgment). *Bell System Technical Journal*, Vol. IX, Oct. 1930, pp. 628–651.	A. E. Bowen C. L. Gilkeson
	"Induction from Street Lighting Circuits—Effects on Telephone Circuits" *A. I. E. E. Transactions*, Sept. 1925, pp. 1065–1071.	R. G. McCurdy
	"Symposium on Coordination of Power and Telephone Plant" *A. I. E. E. Transactions*, Vol. 50, June 1931, pp. 437–478; *Bell System Technical Journal*, Vol. X, April 1931, pp. 151–241.	

Page Number of Statement	*Title*	*Author*
	"Reports of Joint General Committee of National Electric Light Association and Bell Telephone System on Physical Relations between Electric Supply and Signal Systems" Dec. 9, 1922.	
	"Reports of Joint General Committee of American Railway Association and Bell Telephone System on the Inductive Coordination of Railway Electrical Supply Facilities and the Communication Facilities of the Bell System" April 15, 1931; Sept. 1, 1932.	
	"The Design of Transpositions for Parallel Power and Telephone Circuits" *A. I. E. E. Transactions*, Vol. 37, pt. 2, pp. 897–936, 1918.	H. S. Osborne
	"Twenty Years of Transcontinental Service" *Bell Telephone Quarterly*, Vol. XIV, No. 4, Oct. 1935.	J. J. Pilliod

VOLUME XV OCTOBER, 1936 NUMBER 4

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each. The foreign postage is 35 cents per year or 9 cents per copy.

The Bell System Technical Journal

Vol. XV *October, 1936* *No. 4*

Microchemical and Special Methods of Analysis in Communication Research

By BEVERLY L. CLARKE and H. W. HERMANCE

Analysis was beginning to take its place as an important branch of chemistry when, in 1828, Wöhler synthesized urea and the Age of Synthetic Organic Chemistry was born, destined to overshadow analysis for nearly a century. When interest in synthesis began to diminish, in the late 1800's, physical chemistry arose to intrigue the chemical mind. The analyst, thus neglected, had to work with apparatus, techniques and viewpoints evolved for other chemical purposes. In 1910 the Austrian Pregl found it necessary to analyze a sample too small for the then available technique to handle. His solution was the invention of a new kind of analysis—microanalysis, the essential features of which are: reduction of apparatus size and of scale of operations to a point commensurate with sample size; development of entirely new techniques, apparatus and chemical reactions specially suited to analysis; and inculcation in the mind of the analyst of the attitude that analytical problems are, in greater or less degree, research problems, and are to be approached as such, with a mind entirely unrestricted by chemical classicism. This article discusses the applications made by the Bell Telephone Laboratories of microanalytical and related special techniques to communication research and engineering.

THE beginnings of chemistry are lost in antiquity. The basic entities of the early natural philosophers, earth, air, fire and water, gradually gave way to the more numerous and fundamental entities, the elements. In the Middle Ages the alchemists concentrated their talents on an unsuccessful attempt to change base metals into gold. Although these men were, with several notable exceptions, charlatans and fakers, they did focus attention on matter and its objective properties. As early as the 5th century, B.C., Thales of Miletus proposed an atomic theory; but it was John Dalton who twenty-three centuries later formulated the modern Atomic Theory which is the foundation-stone of chemistry.

In the intellectual gropings of man, atoms and molecules, in due course of time, became concepts that explained many phenomena. During all these years there has been a search for the single entity of which all matter was made. Prout's hypothesis of the early 19th century named hydrogen as this single elemental substance. By the end of the 19th century and the beginning of the 20th the one element

became charges of electricity, positive and negative. Today they are the electron and the positron. Out of all the earlier concepts slowly arose the basic idea of chemistry, that the substantial and ponderable part of the world, matter, composed of solids, liquids and gases, is susceptible to controlled transformation.

The devising of tests by which various kinds of matter could be recognized was one of the early accomplishments of the chemist. Many such tests were used by the ancients. It was left for Robert Boyle (17th century), famous for his Gas Law, to conceive identification tests as an important branch of chemistry. Boyle was the first to use the expression "chemical analysis." Lavoisier, of French Revolution era, is credited with having brought about a chemical revolution, one result of which was "quantitative analysis"—methods for determining quantitatively the composition of materials. The Swede Berzelius, working early in the 19th century, analyzed with prodigious industry hundreds of compounds, thus laying the foundation for the quantitative data of chemistry. In the middle 1800's the Belgian chemist Stas repeated and extended Berzelius' work, developing methods and techniques of much greater accuracy.

With the impetus given to it by Stas' work analytical chemistry might have been expected to hold the center of the chemical stage during the 19th century. But in 1828 the German Wöhler synthesized the substance urea from laboratory chemicals. Urea belonged to the vast class of compounds produced by vital processes, and chemists accepted the dogma that these *organic* compounds could not be otherwise produced. Wöhler's synthesis disproved that dogma, and the great Age of Synthetic Organic Chemistry began, destined to occupy chemists' minds for about a century.

Since little attention had been given to analytical chemistry during these years, it became the step-child of the science, useful but not particularly creative. The natural result was that it became a stagnant, static science. It had no special apparatus of its own, but had to be content with the instrumentalities designed for other purposes. Similarly, no one had made any special search for chemical reactions particularly adapted to analysis. Interest in synthesis had also begun to wane. Then physical chemistry burst forth to open up new vistas for the science.

This continued, essentially, until 1910. In the University of Graz, Austria-Hungary, the biochemist Pregl labored for years on a research. Finally he reached a crucial stage of the work. Before him were a few small resultant crystals, whose composition it was necessary to know before further progress was possible. His analyst told him

that the sample was far too small to analyze. So Pregl faced a clean-cut dilemma: either he must start all over again on his research, work for years more on a larger scale, in order to prepare a larger sample, or some way must be discovered to analyze the small sample in hand. Pregl chose the latter alternative, and in so doing initiated an era of development in chemical analysis that has not even yet reached its zenith, namely microchemistry.

Pregl worked in organic chemistry. Simultaneously another Austrian, Emich, approached inorganic analysis from the new point of view. In this country Chamot, at Cornell, concentrated on chemical microscopy. These three, Pregl, Emich, and Chamot, are properly credited with the invention of what has come to be called microanalysis; but their many students and co-workers, as well as scores of independent investigators, did and are doing much brilliant work in the shaping of the science to the practical needs of industry and research.

The basic idea of microanalysis is the reduction in size of analytical apparatus to suit small samples. This has shown the necessity of devising many entirely new methods for carrying out common laboratory operations. Many new chemical reactions have been discovered, on specific search, that have special usefulness in analysis.

The Microanalytical Laboratory at Bell Telephone Laboratories has been established for about seven years. The peculiar nature of many problems arising in communication research and engineering has made necessary the development of many new techniques and types of apparatus. It can be said, in fact, that this laboratory employs a special kind of microanalysis, constituting, for the most part, an original contribution to the science of analysis.

Analysis consists in transforming an unknown material into one or more recognizable substances. These products may then be suitably separated and purified and their quantities measured. Thus weighing, measurement of volume, solution, filtration, washing, evaporation, drying, ignition, distillation, etc., are familiar operations in analysis.

The techniques and apparatus formerly used for these operations, while suitable for the other chemical purposes for which they were designed, were in general ill-adapted to analysis. Apparatus was designed for general utility and manual convenience; that is, to fit the worker's hand rather than the sample. A chemist of the last century would not have thought of using a 1000 cc. beaker to contain 50 cc. of liquid; he would have selected a 100 cc. vessel. But if an analyst had only 0.1 cc. of sample he could find on the shelf no vessel of commensurate size.

Suppose, for example, it is desired to determine traces of heavy metals such as copper and nickel, in a certain plant ash. Because of the disproportionality between the quantities of the elements sought and the size of the apparatus, the errors introduced are large when ordinary methods are employed. Mechanical losses incurred in the many manipulations and transfers of material, over-dilution with resulting incomplete precipitation, contamination both by dust and by substances dissolved from the glass are almost unavoidable. Since the heavy metals represent only a few hundredths of a per cent of the ash, it is obvious that a large sample, perhaps twenty-five grams, is

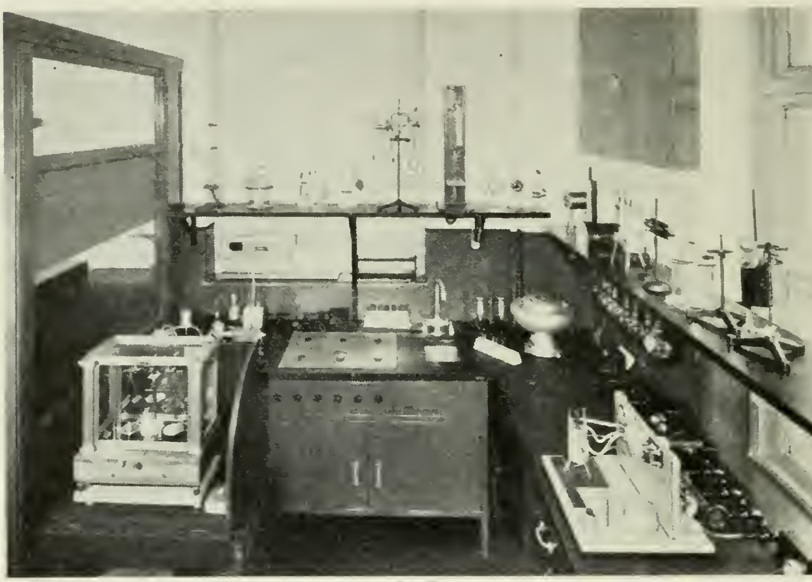


Fig. 1—A corner of the microanalytical laboratory.

required, if an ordinary balance sensitive to 0.1 mg. is used. It would therefore be necessary to start with a kilogram or more of the fresh plant to obtain results which are sufficiently precise.

Other disadvantages of the usual scale of operations are great time consumption, explosion hazards, and costliness of chemicals and apparatus. On a greatly reduced scale, these difficulties frequently tend to vanish. For example, the precipitation of the sulphides of heavy metals is ordinarily avoided wherever possible in quantitative analysis. This is because they are slimy and difficult to filter as ordinarily precipitated with hydrogen sulphide. To filter and completely wash a gram of lead sulphide might be a matter of several

hours and there is great danger that the precipitate will occlude other constituents of the solution. The long exposure on the filter increases the danger of oxidizing the precipitate. On the other hand, a few milligrams of lead sulphide can be precipitated in a sealed tube by a process which produces hydrogen sulphide up to a pressure of ten atmospheres, yet at very little risk of explosion. The resulting precipitate is granular, can be filtered in one or two minutes and its tendency to occlude other metals is much smaller.

Until recently, the most important industrial applications of analysis were the evaluation of raw and finished products and the control of manufacturing processes, neither of which often demanded special technique. With the growing technical trend in commercial production, however, the analyst is being called upon to provide new services. Industrial research must be guided by frequent analyses, both to determine the nature of newly formed products and to gain knowledge of the mechanisms of particular processes. The great diversity of materials, natural and synthetic, and the intricate and often delicate mechanisms embodied in devices of modern manufacture, have enormously enlarged the problem of tracing the causes of failure both in the finished product and in the processes entering into its production. Trouble often arises from obscure defects in materials, the nature of which must be discovered by analytical studies. Impurities, minute foreign inclusions, corrosion and tarnish films, chemical changes occurring with aging or produced as an inherent result of the particular combination of materials used, may contribute.

To make effective use of chemical analysis either as an industrial research tool or as a means of diagnosing manufacturing and maintenance difficulties, the necessity of improving the technique is beginning to be recognized here as in other fields. Great flexibility is needed to fit the operations to highly specific problems. Ability to handle and observe small quantities is frequently necessary because of the minuteness of the phenomena in question. Rapidity is often essential because of the possibility of tying up production, pending solution of the difficulty.

Realization of conventional limitations has stimulated the search for new methods of approach by which the refinement and extension of analytical technique might be accomplished to fulfill the special needs of both science and industry. An outstanding result has been the development in the analyst of a new mental attitude. He seeks to attain his goal first by reducing the scale of operations to a degree consistent with the small quantities of material frequently handled;

second, by augmenting his ability to make observations through the use of adequate instruments; and third, by employing specific and highly sensitive reactions as well as conversion products of high molecular weight.

Micromethods serve the obvious purpose of analyzing minute amounts of material, thereby providing information otherwise unobtainable. Actual experience at Bell Laboratories, however, has shown that the reduction in magnitude of operations frequently permits analyses to be carried out with greater rapidity and more certain results even when the quantity of sample available is not a considera-

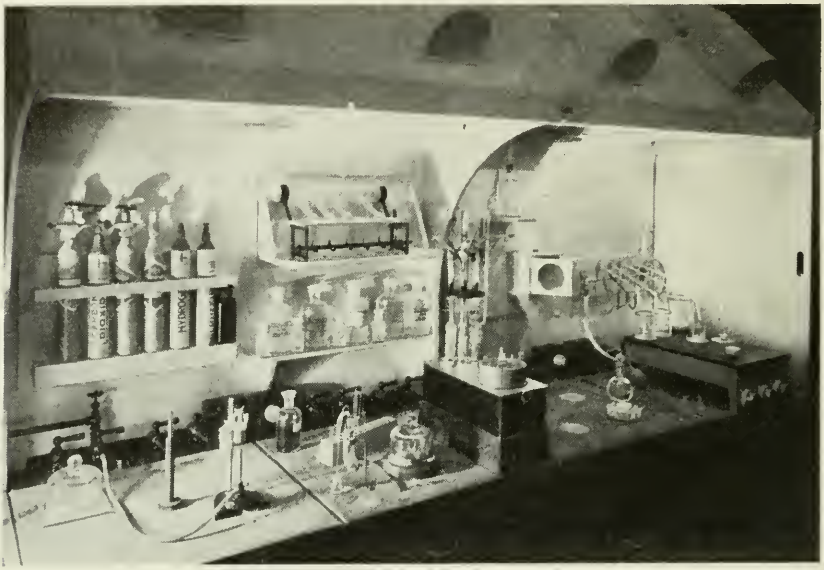


Fig. 2—General view of hood in the microanalytical laboratory, giving some idea of the relative size of glassware and other equipment used.

tion. The construction of apparatus, when more intricate set-ups are necessary, is far easier and more economical on a small scale and the breakage is less. Reactions run their courses more quickly and are more easily controlled. Reagents may be used whose costliness would be prohibitive on a larger scale. These advantages, added to the capacity to make minute observations, provide a technique of great flexibility, a fact repeatedly demonstrated by successful applications to problems arising in the design, manufacture and maintenance of telephone equipment.

In microqualitative examinations, an effort is usually made to bring the unknown material into solution in a volume not exceeding a

few tenths of a cubic centimeter. Identification reactions are then carried out in single small drops of the solution. To accomplish this, a number of procedures are available. One which is very generally applicable, yielding information of a particularly specific and positive character, is that of carrying out the reaction directly under the microscope. By this means, quantities of elements ranging between a thousandth and a few hundredths of a milligram may be detected. The drop to be examined is placed on a glass slide, the reagent introduced from a capillary pipette and the progress of the transformation watched under magnifications of between fifty and three hundred diameters.

In addition to revealing the presence of minute amounts of reaction products separating from the solution, the use of the microscope facilitates study of the individual particles composing such precipitates. A number of properties are thereby brought into analytical significance which might not otherwise be observed or utilized. In ordinary practice, the analyst is guided in his conclusions as to the presence of an element simply by the bulk formation of a precipitate, specific recognition of which is based only on characteristics readily apparent to the unaided eye, such as color and gross structure. Under the microscope, however, the crystal structure and similar distinctive morphological features, color, transparency, index of refraction, behavior toward polarized light, characteristics of growth and other specific properties may all be studied and employed to give greater certainty to the identification. Further, it is frequently possible to identify the individual components of a mixed precipitate, thereby obviating the necessity of separation. Thus, the double salt potassium mercuric thiocyanate gives insoluble compounds with a large number of the bivalent metals. To apply this reagent to a solution containing several metals would result in the formation of a precipitate which to the unaided eye would yield very little specific information. Under the microscope, the experienced analyst, in a single observation, can often tell from the known habits of the crystals produced by various combinations of metals, the nature of the mixture.

Recognition of substances is not always confined to the precipitation of insoluble reaction products, but soluble salts, when the solution is carefully evaporated, sometimes possess sufficiently distinctive morphology to permit direct identification. In this way minute amounts of sodium chloride have been detected in dust deposits collected near the sea-coast. In the identification of traces of organic material, valuable information is frequently obtained from microscopic studies of the crystalline deposits produced when the substance is

sublimed directly on the slide. Crystallographic constants and other optical properties as well as the melting point and solubility may be readily determined on the sublimate even though only a fraction of a milligram is available.

Another advantage to be gained by reactions carried out under the microscope is that of performing tests *in situ.* In practice, it is sometimes difficult to isolate physically the particles of material to be studied. For example, it may be desired to learn something of the nature of a tiny inclusion embedded in a metal or a thin film of corrosion product present on its surface. Here again, the microscope is of great service, both in guiding the physical manipulations necessary to restrict the action of the reagents to a localized area and in observing the actual identification reaction (usually chosen to yield an intensely colored product or bubbles of gas, rather than a precipitate).

When, instead of the commonly used qualitative reagents, compounds are employed which are capable of more specific and sensitive reactions and which yield intensely colored products rather than precipitates, such products may be instantly recognized, even in a single drop of solution. Within recent years, many organic compounds have been developed for this purpose. The use of these, and of a number of inorganic compounds giving highly characteristic color reactions, constitute the basis for a new technique combining rapidity, simplicity, and certainty of identification without the use of the microscope. The drop to be examined is placed on a white background such as a porcelain plate. A drop of the reagent is added and the color change, which may involve a sequence of changing shades, is observed. In cases where turbidity is also significant, a black porcelain plate is used. Because of the highly specific nature of these reagents, it is often unnecessary to resort to a preliminary group separation. Because of its simplicity, the technique is particularly useful for field investigations.

A modification of the drop analysis procedure described above consists in bringing the drop under examination together with the reagent onto loose-textured paper such as filter paper. The colored product which forms is adsorbed on the fibers of the paper at the center of the drop, while the solution spreads out due to capillarity. The capillary action of the paper fibers is sometimes utilized to devise rapid separations, thus enabling the analyst to detect two or more elements simultaneously. In a typical case, a solution contains copper and nickel. A drop of this solution, acidified with acetic acid, is brought onto the paper which is impregnated with hydrorubeanic acid. The bluish-black copper compound is insoluble and therefore

forms first, near the center of the drop. The nickel compound, being more soluble, does not form until the solution has spread out to a considerable extent and most of the acid has evaporated, when it is visible as a violet zone near the periphery of the drop.

Papers which have been impregnated with specific reagents and preserved in the dry condition are extremely useful both in the laboratory and the field. In the analysis of gaseous substances or materials readily volatilized, these dry test papers have been very satisfactory. Arsine, stibine, hydrogen sulphide, sulphur dioxide, hydrocyanic acid and other objectionable gases present in minute quantities in the atmosphere, may be detected and their approximate concentrations determined by passing a stream of the air through the fibers of a suitable test paper.

In order to obtain solutions to which identification tests may be applied, some preparatory chemical treatment is necessary. The initial solution and concentration of the sample, its recovery from inert material as well as its separation into convenient analytical groups require laboratory operations capable of dealing with a few drops of liquid and often with a fraction of a milligram of solid. A variety of types of apparatus and special processes have been developed to facilitate these operations. In a few cases, reduction in size has alone sufficed; more often such reduction, with retention of the original form of the apparatus, results unsatisfactorily and new principles must therefore be followed in the microdesign.

For example, it is obvious that the usual folded paper cone cannot be satisfactorily reduced in size to permit the removal and recovery of suspended matter from a few drops of liquid. Microfiltration may be accomplished in a number of ways but the most convenient is to draw the liquid into a capillary pipette provided with a retaining well which holds a very small pellet of the filtering medium. In this way a single drop may be filtered and completely washed in a few seconds, the residue being concentrated in the tiny filter plug from which it may readily be redissolved in a trace of acid. Practically all of the common analytical operations have been reduced to a microscale and may be carried out with rapidity and precision in equipment of appropriate design.

It may be of interest to note here a few of the more ingenious devices and processes that have been developed to aid qualitative microanalysis. The electrolytic cell of H. Brenneis provides for the precisely controlled electrolysis of a drop of solution, at the same time permitting continuous observation of the electrode surfaces under the microscope. By its use, 0.001 mgm. of copper may readily be recog-

nized and as little as 0.0001 mgm. of zinc has been observed on a previously coppered cathode. The electrodes are formed by encasing closely spaced platinum wires in glass which is subsequently cut and polished through a perpendicular plane, thereby exposing cross-sectional areas of the wires. The drop to be electrolyzed is placed on the polished glass surface and the portion covering the platinum areas observed under the microscope.

A process rarely used in ordinary analysis but of great service in microwork is that of sublimation. A few thousandths of a milligram of a volatile crystalline solid may be separated in this way from a

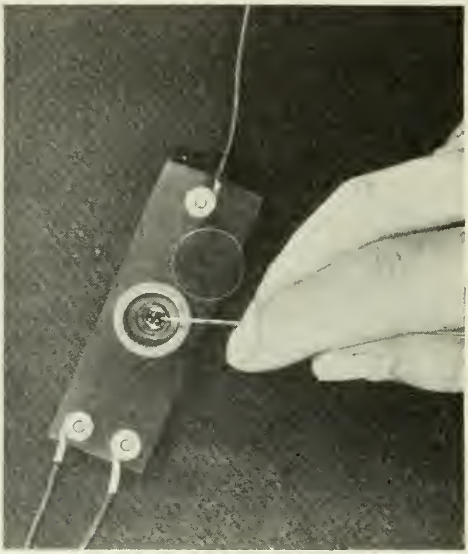


Fig. 3—The electrolysis cell shown here, when used under the microscope, permits one to observe the deposition of metals from tiny drops of solutions. Less than a thousandth of a milligram of copper or zinc may be detected by its use.

large bulk of inert material in a condition that permits immediate treatment with reagents. In order to apply this process to dusts, corrosion products and other frequently-encountered materials, an improved microsublimation chamber has been designed by this Laboratory. The apparatus is so arranged that both temperature and pressure can be regulated. The vapor condenses on a water-cooled microscope cover-glass and the recovery is practically quantitative. As little as 0.002 mgm. of mercuric iodide was found to give a deposit of definitely recognizable crystals.

A phenomenon that has long been familiar, yet not applied to analysis until its value in microwork was recently demonstrated, is

the production of "schleiren" or refraction lines when two fluids of differing optical density are added together without mixing. One liquid is contained in a flat optical cell of a few tenths cc. capacity. The other is slowly introduced below the surface of the first liquid from a capillary orifice. The refraction effects set up are observed through a horizontally mounted microscope under controlled illumination. When one of the liquids is known, such observations are the basis for both specific gravity and refractive index determinations. So sensitive is the method that a difference of 0.0001 in the refractive indices of the two liquids is still detectable. It therefore affords an excellent test for purity. If, for instance, two fractions of a distilled liquid are added together with the production of "schleiren," it may be assumed that the original liquid was not a pure substance.

Glass capillary tubes have shown great versatility in microwork. When the quantity of material operated upon is exceptionally small, as, let us say, in the case of a foreign deposit on relay contact points, their use affords distinct advantages. Almost every operation can be executed through appropriate adaptations of capillary technique. Thus reactions may be carried out under pressure in sealed capillaries when the volume of liquid is only a few thousandths cc. and the whole process watched under the microscope. Distillation and sublimation are processes to which capillaries are especially suited. Suspensions may be centrifuged or filtered in capillaries. In the latter case capillary attraction is the force that draws the liquid through the filtering medium.

The possibilities of capillaries may be illustrated by a practical example. Tiny discolorations were found on the surface of a polished silver sheet used in photocell manufacture. Mercury contamination was suspected and an analytical confirmation was desired. The procedure was as follows: A 1-mm. capillary tube was drawn out to form a pipette having a very fine tip. With this a very small drop of nitric acid was transferred to one of the discolored areas under the microscope, and allowed to act for a few seconds, after which it was removed, transferred to a capsule and the excess acid evaporated. The residue was re-dissolved in a small drop of water and drawn up into a second capillary tube containing a few mm. of No. 32 copper wire. Both ends were sealed and the tube heated in boiling water for a few minutes, after which one end was opened and the liquid withdrawn by means of a finer capillary. The open end was then drawn out to a very fine tube of microscopic bore. The closed end was heated by a microflame, gently at first to drive out moisture, then strongly until the glass had completely fused about the copper wire. Heating in

this manner was continued almost to the constricted portion. When the capillary, after cooling, was examined under the microscope minute globules of condensed mercury were plainly visible. In this process, the mercury is displaced from solution by the copper and

Fig. 4—With this electrolytic cell as little as a milligram of various heavy metals may be precisely determined.

deposits on the wire from which it is subsequently distilled. As little as 0.001 mgm. can be readily detected in this way.

A considerable part of the microanalyst's task is the physical isolation and recovery of the material on which his analytical operations are to be performed. His problems very often require examinations of minute particles or aggregates of foreign substances which have become attached to or embedded in the surface of a material. He may also be required to isolate and study the structural units which compose a given formation. For example, a deposit occurs on the surface of a metal as a result of corrosion. This deposit is not of a homogeneous nature but is built up in successive layers, each of which differs in composition. To obtain a satisfactory picture of the mechanism of the production of the deposit it is necessary to know the composition of each separate layer.

The mechanical manipulations necessary to obtain sample material frequently tax the analyst's ingenuity more than the analysis itself. The work is usually carried out under the low powers of a microscope, preferably of the binocular type. Much of the technique of the biologist has been appropriated by the microanalyst in this phase of his work. Various types of dissecting tools find ready application here. The dental engine with its various attachments, such as drills, burrs, carborundum wheels, has been found extremely useful for drilling out inclusions in metals and for the removal of hard surface films. The micromanipulator, an instrument originally designed by biologists to perform intracellular operations, has recently been employed in microchemical work with very satisfactory results. This instrument furnishes the means of regulating with great precision the movement of needles, capillary pipettes, electrodes, electrically heated platinum wires, etc., under relatively high magnifications. It offers great promise where particles of exceptionally small dimensions are to be studied.

A number of methods have been developed and used by the Laboratories' Microchemical Group for the collection and study of central office dusts. A device which deserves particular mention is the impinger, an adaptation of which has been employed to remove dusts from the extremely localized area represented by a single relay contact point. The device is so constructed that the particles, after being picked up by suction, are projected at high velocity against a microscope slide, the surface of which is coated with an adhesive medium. The slide is removed from the apparatus and the dust subjected to physical and chemical treatment to determine its nature.

In quantitative microanalysis, the analyst is faced with the added

problem of weighing a few milligrams of material with the same precision as might be obtained with samples of the usual size. The Nernst quartz fiber balance, capable of weighing to a few ten-thousandths of a

Fig. 5—The micromanipulator finds use when it is necessary to operate on unusually small particles or within areas bounded by the microscopic field.

milligram, has proved very useful in certain types of work, where the total load does not exceed a few tenths of a gram. Its application to chemical analysis, however, is quite limited because of the difficulty of reducing the load to this extent. The quantitative extension of

microtechnique consequently made little progress until a beam balance was produced by W. Kuhlmann of Hamburg, which was capable of weighing to a thousandth of a milligram with no appreciable change in sensitivity with loads up to 20 grams.

With the perfection of this essential instrument quantitative microtechnique developed rapidly, and because of the economy of time and material it is in many cases actually displacing older methods operating on the usual scale. This is particularly true in organic analysis where the methods are ordinarily tedious and expensive. Pregl, who received the Nobel prize in 1923, worked out rapid, precise micromethods for the determination of carbon, hydrogen, nitrogen and various organic radicles, which require only a few milligrams of sample. As an example of the great practical value of such methods may be cited an instance mentioned by Cornwell [1] in which a complex organic compound was synthesized with a yield of about a gram of product. The labor and material involved brought its cost to about $5,000. An analysis was required as a check on the composition, and by the usual methods this would have required 0.2 gram of sample at a cost of $1,000. The analysis was actually made by the micromethod on 2 milligrams of sample, cost $10.

Most of the general equipment originally devised to facilitate micro-operations in qualitative analysis is also applicable to quantitative work. Considerable additional equipment is required, however, for the recovery, conditioning and quantitative measurement of the final transformation products of the analytical process. Precipitates are collected and weighed either by centrifuging in suitably shaped vessels or on suction filters which are essentially miniature reproductions of those used in ordinary work. An innovation in filtration practice consists in the use of an inverted filter which is weighed together with the microbeaker in which the final precipitation takes place. The clear liquid and washings are simply drawn off through the filter. This obviates the necessity of completely transferring the precipitate to the filter, thereby avoiding losses that are otherwise almost certain. Duralumin blocks of various designs have been found excellent for drying and conditioning precipitates. The high heat capacity afforded by the large mass of metal insures a very constant temperature. Various types of micromuffle and combustion furnaces have been devised. Their small size greatly reduces construction costs and permits a more generous use of quartz or platinum linings.

For the measurement of liquids, microburettes are available which can be read to 0.001 cc. When the quantities are too small for

[1] Cornwell, R. T., *J. Chem. Education*, *5*, 1099–1108 (1928).

measurement by micromodifications of the conventional gravimetric and volumetric methods, the microscope sometimes may be ingeniously applied to quantitative determinations. Colorimetric and turbidimetric measurements may thus be made on a few thousandths of a cubic centimeter of solution contained in a capillary tube. The relative quantities of two or more components of a mixture may frequently be approximated from area determinations made on the individual particles in the microscope field. Instead of weighing the mercury recovered by capillary distillation, the condensed globules may be united by centrifuging and the mass of the resulting single globule estimated from diameter measurements under the microscope. Similarly the analysis of a minute amount of gas may be carried out by measuring the shrinkage in diameter of a single microscopic bubble as the absorption reagents in which it is immersed are changed.

A striking example of a quantitative determination so contrived that the final measurement is performed microscopically is the micro-molecular-weight method of Barger. Two solutions, one known and the other containing the unknown, are placed in a capillary with a small air bubble separating them. The ends of the capillary are sealed and the lengths of the two liquid columns measured on a micrometer scale. After several hours the measurement is again made, and repeated at intervals until the column lengths become constant. When this occurs, the vapor pressure of the two solutions will be identical and since vapor pressure is a function of the molar concentration, the latter may also be assumed to be the same in each solution. Knowing the original weight concentrations and the molecular weight of one of the substances, that of the other may be calculated from the change in the volumes of the two solutions necessary to bring about equilibrium.

The application of quantitative microtechnique to engineering and research problems at Bell Telephone Laboratories has required a considerable amount of development work directed toward the improvement of apparatus and the creation of new types of technique. Thus, in order to carry out micrometallurgical analyses, several forms of electrolysis cells were developed which permit the determination of metals such as copper, zinc, nickel, lead, cadmium, tin and others. With these cells, using five milligram samples, the same accuracy is attained as in ordinary analysis on half a gram. One of these cells, designed for the analysis of extremely dilute solutions, permits the qualitative detection of one part of copper, zinc, or lead in 100,000,000 parts of water and is particularly useful in isolating minute quantities

of heavy metal impurities in such metals as aluminum or nickel, or in examining waters after use in corrosion experiments.

In order to reduce the errors of weighing in microgravimetric analysis, it is obviously desirable to obtain transformation products having the greatest possible mass. To facilitate manipulation, it is also desirable to deal only with products of a coarse crystalline nature

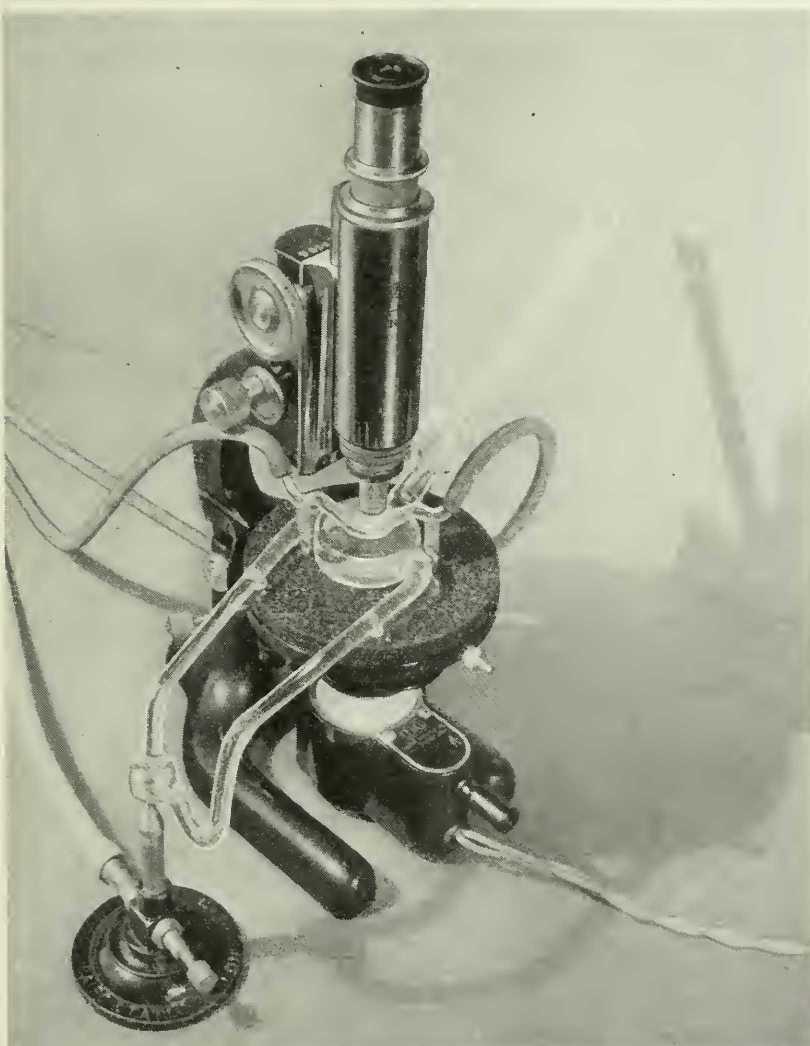


Fig. 6—The glass hot stage assembly shown here provides for the determination of melting points or the sublimation of volatile substances under the microscope.

which may be readily centrifuged, filtered and washed, and which exhibit a minimum tendency to adhere to the wall of the containing vessel. The simple insoluble salts, oxides or hydroxides usually obtained in the older, classical methods, rarely meet these requirements satisfactorily and recently much attention has been given the search for more suitable quantitative precipitants. The result has been an increased use in microwork of organic, complex and double salts of high molecular weight. Thus sodium is no longer weighed as the sulphate, which provides only a threefold increase in the weight of the sodium present, but rather as crystalline sodium zinc uranyl acetate having a molecular weight of 1591 or a weight equal to 69 times that of the sodium present. Silver may be collected and weighed as silver-copper propylenediamine iodide of molecular weight 939. Aluminum, instead of being precipitated as gelatinous aluminum hydroxide, which is difficult to filter, may be separated and weighed as the crystalline oxy-quinolate with an eighteen-fold increase in weight.

Another possibility of chemically amplifying weight consists of employing a train of reactions the final weighable product of which has a high molecular weight. Although it may not contain the element originally sought, this product is still stoichiometrically related to it and provides the basis for quantitative estimation.

Although Pregl, Emich and Chamot are considered the founders of modern microchemistry, there were a number of earlier isolated instances of such methods being used. The earliest attempt to organize qualitative microchemical methods systematically was made by Boricky in 1877 who applied the technique to petrographic studies and wrote a treatise entitled "Elements of a New Microchemical Analysis of Minerals and Stones." In 1885, Haushofer in his book "Mikroskopische Reaktionen" provided a rather complete description of reactions carried out under the microscope, his work covering most of the common elements. It remained for Professor H. Behrens of Delft, Holland, and H. Schoorl to expand the technique to a point where minute quantities of substances could actually be separated and manipulated to permit the application of common analytical operations. In the biological field, H. Molisch developed the technique of applying microreactions directly to plant tissues and in this way was able to identify many intracellular substances. His book "Mikrochemie der Pflanze," and that of Mayerhofer, "Mikrochemie der Arzneimittel und Gifte," are well known.

In Europe, particularly in Germany and Austria, awakening interest in the application of microchemical methods to industrial problems is evidenced by the appearance of a large number of articles on the

subject in current journals. Until very recently, however, American chemists did not appear to have fully realized the technical possibilities and advantages of micromethods. The installation of a completely equipped microchemical laboratory at Bell Telephone Laboratories about seven years ago, constituted, the authors believe, a pioneer step in the direct application of microtechnique in its broadest sense to engineering problems. Since that time, a number of industries and industrial institutions have made similar installations with favorable reports as to their usefulness.

In order to convey a more concrete idea of the types of problems in which the microchemical approach has been particularly helpful at Bell Laboratories, this paper will be concluded with a few actual examples.

The equipment used in the telephone plant and associated industries contains numerous small functional parts, concerning which analytical information is frequently needed. Such information may be desired in connection with laboratory studies required in the design of the apparatus or it may be needed because of unsatisfactory performance in service, traceable to some irregularity in the particular part. Obviously such results as might be obtained by compositing the large number of parts necessary for an ordinary analysis would be inadequate. Both the peculiarities of the individual case and the variations in quality and composition will be obscured unless the analytical study be made to include only particular specimens. Thus single relay contact points, weighing between five and ten milligrams, have been quantitatively analyzed to check the composition of the alloy when excessive deterioration was noticed. The gold plating, amounting to a few tenths of a milligram, has been precisely determined on small areas of handset transmitter parts for the purpose of observing the uniformity of the coating.

In studying the various phenomena occurring in vacuum tubes and photocells the Microchemical Laboratory has frequently been requested to identify and occasionally to analyze quantitatively various metallic films and surface deposits in which the total material has ranged from a few thousandths to one or two milligrams. The distribution of caesium on the various surfaces of the photocell was quantitatively studied by microanalysis, the greatest quantity of caesium present in any one determination being about 1.5 mg. Single filament wires weighing from eight to thirty milligrams have been analyzed, the thoria determined in the case of tungsten filaments, while nickel wires were examined for copper, iron, silicon and manganese.

Micromethods have been utilized in studying variations in composi-

tion between very thin layers of a material. In a typical case, sheets of an iron-cobalt magnetic alloy were observed to undergo a change in properties after rolling to diaphragm thickness. It was suspected that the surface had lost iron through oxidation and subsequent mechanical removal. By using pure silica abrasive, a layer of metal was removed from the surface corresponding to a thickness of about 0.005 millimeter. The metal was then extracted from the abrasive by acid treatment and the iron-cobalt ratio determined microchemically. A similar technique has been applied to tinned copper wire, to determine the quantity of copper dissolved by tin during the tinning process and the extent of its migration to the surface of the coating. It has also been used to study the alloy layer formed between the zinc coating and the iron base in sherardizing and galvanizing processes. Micromethods applied to thin films have been used in the study of the copper-oxygen ratio variation in copper oxide rectifier discs.

The inclusion of foreign particles in the surfaces of metals and other materials occasionally occurs as a result of faulty conditions of manufacture. Determination of their nature and source is naturally necessary before remedial measures can be taken. Since the particles are small, frequently even of microscopic dimensions, microchemical methods in such cases are the only ones practicable. Thus small hard particles were observed in an experimental lead cable alloy which were at first thought to consist of segregated impurities. Microanalysis showed these particles to be composed of iron and nickel, indicating that they had probably been accidentally introduced into the surface during the extrusion process. Paper removed from a condenser that had failed on test was found to contain microscopic particles of iron, iron rust, brass and carbonaceous aggregates which had also been accidentally incorporated during manufacture.

The isolation, detection and determination of traces of impurities or substances otherwise associated with large quantities of a given material have benefited through the application of microtechnique. Methods have been worked out for determining sulphur and phosphorus in steel and for silver in lead in which these impurities are all present in quantities less than 0.001 per cent. Acetic acid has been isolated from the corrosion products occurring when lead cable is exposed in creosoted wood ducts. The quantity actually present is usually very small, amounting to between 0.01 per cent and 0.001 per cent of the weight of the corrosion product.

Electrolytic corrosion in the windings of relays and other similar types of apparatus may occur as the result of minute quantities of salts present in the insulating materials or acquired from the manu-

facturing environment. These salts furnish electrolytes when the humidity is relatively high, dissolving the copper at certain poorly insulated points in the winding until the wire is eventually severed. In tracing the sources of this type of corrosion, microtechnique has afforded the only satisfactory method of attack.

Recently studies have been made to learn the effects of the dusts found in the telephone central office on the functioning of machine switching equipment. In these studies, microchemical methods have figured largely in the identification of the individual dust particles as well as in the quantitative determination of the major type of components.

Switchboards and Signaling Facilities of the Teletypewriter Exchange System *

By A. D. KNOWLTON, G. A. LOCKE and F. J. SINGER

The development of a nationwide teletypewriter exchange system in the United States required the design of switchboards and signaling facilities adapted to this special service. The two types of switchboard now in use are described in this paper, and the operation of the circuits by means of which connections between the various subscribers are established and supervised by the operators is explained.

A NATIONWIDE teletypewriter service giving direct connection between subscribers for the exchange of written messages by means of the teletypewriter in a manner similar to the service offered by the telephone system for the exchange of spoken messages was offered to the public as a new aid to business by the Bell System on November 21, 1931. This service, known as the teletypewriter exchange (*TWX*) service, introduced a switching technique which, although familiar in the telephone art, involved many new technical problems when applied to the telegraph art.

Records show that during the nineteenth century some telegraph exchanges were established at which connections could be made on a message basis for to and fro telegraph communications between subscribers. These earlier exchanges had a commercial appeal although the various forms of subscriber instruments then used were slow and required considerable skill for operation. Later, when the telephone was introduced, these exchanges gradually disappeared because the public naturally preferred the more convenient instrument. With the introduction of the modern teletypewriter the telegraph exchange idea was again revived because the teletypewriter, being very similar to an ordinary typewriter and permitting an accurate written record of a to and fro communication, has, from a subscriber standpoint, overcome the objectional features of the early telegraph instruments.

The private line telegraph and teletypewriter service furnished by the Bell System has formed a very important background for the new teletypewriter exchange service. The older service, which provides relatively permanent networks interconnecting various stations in a predetermined manner for a predetermined time, has been available to the public since about 1890. During the earlier period it was used

* Published in *Electrical Engineering*, September, 1936. Presented at A.I.E.E. Southwest District meeting, Dallas, Texas, October 26–28, 1936.

chiefly by the press and brokers and was operated on a Morse telegraph basis, generally using composite or simplex line facilities. Later, with the introduction of the modern teletypewriter, the carrier telegraph, and other improvements, together with the growth of American business and the demand for rapid and accurate written communications, this private line business expanded rapidly and service was furnished not only to the press and brokers but also to other financial institutions, manufacturers, government bureaus, police departments, and a wide variety of retailers and distributors of goods. This business has become nationwide. Many of these private line systems are provided with switching facilities for use by the customer in each system, although the supervisory arrangements are rather elementary.

In addition to the private line telegraph service and the arrangements which had been developed and applied to that service, the many developments in the telephone field formed an important contribution to the teletypewriter exchange service. It is obvious that in providing *TWX* service, which is a point-to-point service with connections set up and taken down on the subscriber's order, use can be made of many traffic and service practices used in the telephone service. Furthermore, certain telephone apparatus such as switching relays, cords, plugs, etc., can be employed to advantage.

With this background, when it was decided to furnish a nationwide teletypewriter exchange service to the public, Bell System engineers had the problem of determining what general plan of design to adopt. There were two alternatives: (1) to provide a service using the telephone plant and existing telephone switchboards, or (2) to provide separate switchboards for use with the telegraph plant. The important advantages of the first plan are:

(*a*) The switchboards and signaling arrangements designed for and in use in the telephone plant could be employed.

(*b*) The same operating groups handling the telephone service would handle this service. Inasmuch as telephone service is on a 24-hour basis throughout the country, the *TWX* service could be furnished on the same basis with a relatively low operating cost.

The disadvantages of the first plan are:

(*a*) Because the teletypewriter operates on a d-c. basis it would be necessary to provide an oscillator and associated apparatus at the station to generate an audio-frequency alternating current for modulation by the signals sent by the teletypewriter, and a rectifier to convert the a-c. pulses received from the distant station to d-c. pulses for operation of the receiving mechanism of the station teletypewriter. Furthermore, it would be necessary to furnish a telephone instrument at the station to permit the subscriber to communicate with the operator unless a teletypewriter or other type of recording instrument were provided at each operating position.

(*b*) Relatively expensive telephone lines known as inter-toll trunks would be required between central offices. If the cheaper telegraph channels were used as

inter-toll trunks it would be necessary to provide frequency converters at each terminal to translate the frequency band required on the subscriber loop to a band suitable for application to the telegraph inter-toll trunks. If the telegraph channels were used between switchboards it would also be necessary to provide the operators with teletypewriters or other means of communication because the telegraph channels do not permit oral communication.

(*c*) A number of miscellaneous engineering and plant problems other than those listed in (*a*) and (*b*) would be introduced if standard telephone facilities were used to interconnect the stations in the teletypewriter exchange network.

After due consideration of all these factors it was decided to utilize the telegraph plant and to design and provide the necessary teletypewriter switchboards and inter-office signaling arrangements. By following this plan it has been possible to establish service on a nation-wide basis using switchboards at the larger switching centers and employing modified telegraph private wire testboards at the smaller centers.

This paper describes the signaling and switching arrangements used in the present system, and particularly the two principal types of switchboards that are in use. The discussion is limited to the most important signaling and switching arrangements, as the transmission features are described in another paper.[1] A description is included of the principal factors entering into the design of the more important circuits used in these switchboards: the subscriber lines, inter-toll trunks, and cords. The subscriber line treatment is divided into three broad classes: local subscribers having either attended-only or unattended service; distant subscribers served over telegraph toll line facilities; and distant subscribers served over telephone facilities. Particular attention is given to the fundamental problem of providing supervisory signals over the telegraph lines used as inter-toll trunks in the inter-office connections.

Teletypewriter Switchboards

To reach subscribers in all parts of the country there has been established a network of teletypewriter switching points interconnected by telegraph lines. At each of the larger switching points a teletypewriter switchboard is provided, the principal switchboards being the No. 1 Teletypewriter Switchboard having a capacity of 3,600 subscriber lines, and the No. 3*A* Teletypewriter Switchboard having a capacity of 1,200 subscriber lines. The former, a general view of which is shown in Fig. 1, is used in large cities such as New York and Chicago, while the latter, a general view of which is shown in Fig. 6, is used in smaller cities such as Pittsburgh and Kansas City.

[1] "A Transmission System for Teletypewriter Exchange Service," R. E. Pierce and E. W. Bemis, this issue of the *Bell System Technical Journal*, and *Electrical Engineering*, v. 55, September 1936, pp. 961–70.

Fundamentally, a manual switchboard consists of two parts: the terminations for subscribers lines and inter-toll trunks, and the switching facilities used by the operators in interconnecting the lines and trunks. The line and trunk terminations are in the form of multiple jacks and lamps located in the jack field and are accessible to all operators. The switching facilities, or cords, together with the means for communication to subscribers or other operators, are individual to each operator and are, in general, located at the keyshelf. Although the design of the switching equipment and the multiple are to some

Fig. 1—No. 1 Teletypewriter Switchboard at New York, N. Y.

extent dependent upon each other, the principal factors influencing the design are, for the purpose of discussion, considered independently.

No. 1 Teletypewriter Switchboard Position Equipment

The No. 1 Teletypewriter Switchboard position consists essentially of a teletypewriter for the operator's use in sending and receiving the instructions for establishing the connections, together with a number of cords for making the various interconnections between the line terminations. The number of these cords necessary for the efficient functioning of an operator is the most important factor governing the width of the position, a primary consideration in the design of a switchboard.

The number of cords per operator is dependent on the average time required to set up and disconnect each call (known as the average work time per call) and the average communication time per call. Whereas the former can be forecast quite accurately by the operating characteristics of the circuits, the latter is dependent on the commercial application of the service. To insure the provision of an adequate number of cords it was necessary to allow for the longest average communication time which could be reasonably anticipated. The analysis of the average work time per call together with the forecast communication time resulted in the requirement being set up for a maximum of 18 cords per operator.

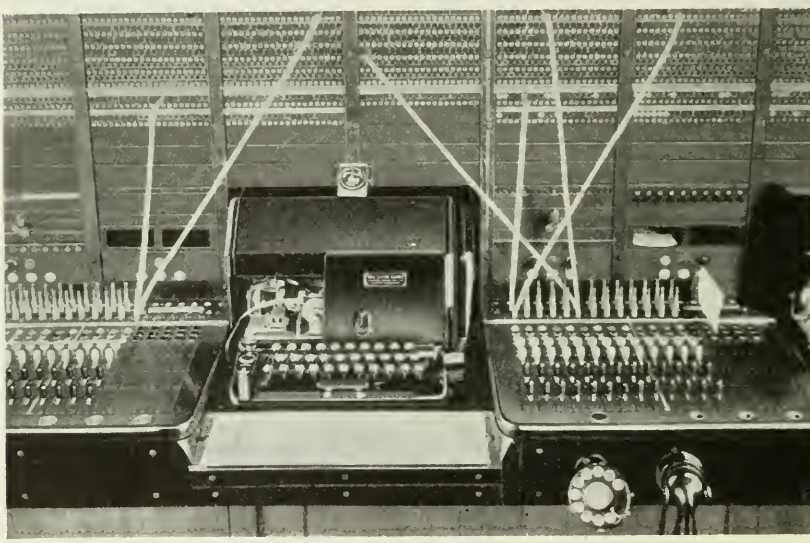


Fig. 2—Keyshelf arrangement of No. 1 Teletypewriter Switchboard.

With the requirement for the position equipment established at 18 cords (and one teletypewriter for communication purposes), the width of the position was determined to be approximately 34 inches, or the width of four panels of the jack field, each panel being $8\frac{1}{2}$ inches in width. The division into an even multiple of panels is for constructional purposes, to separate the switchboard into sections for manufacturing. It was, however, necessary to adopt a new type of keyshelf construction, shown in Fig. 2, to provide for the operator's teletypewriter.

Because of its large size, the teletypewriter was located as low as possible to minimize blocking the jack field. This required the pro-

vision of a teletypewriter shelf the lower edge of which was at the same height as the lower edge of the adjacent keyshelves. The depth of the teletypewriter made it necessary to recess it in the jack field. This recess was obtained by cutting off one stile strip and adding a longitudinal detail the entire width of the section to support the lower end of the cut stile strip. The teletypewriter shelf was placed on rollers to permit its sliding out easily for maintenance accessibility.

The teletypewriter, being the operating center of the position, has nine cords on each side to locate all 18 cords within easy reach of the operator. Because of this arrangement it was not possible to make the position boundaries coincide with the section boundaries as this would require two keyshelves of nine cords each per section with the consequent waste of equipment space for the supports between adjacent keyshelves. This loss of space was reduced by providing one 18-cord keyshelf per 2-position section and associating one half of the cords with the teletypewriter to the left and the other half with the teletypewriter to the right. This caused an overlap of the position and section boundaries so that the nine cords on the left end of each section form a part of the right position of the adjacent section to the left.

No. 1 Switchboard Multiple Equipment

The primary objective in the design of multiple equipment is the provision of line terminations in a form that will make each line readily accessible to every operator, taking into consideration the physical limitations imposed by the operator's reach. Previous experience in the design of telephone switchboards has determined that, for a subscriber switchboard, satisfactory operating conditions may be obtained in respect to the horizontal reach of the operator by the multipling of the line terminations on an 8-panel basis (using the standard $8\frac{1}{2}$ inch panel) giving a distance of 68 inches from one appearance of a line to the next. The maximum reach in each horizontal direction will then be half of this distance, or 34 inches. This was, however, reduced to a 6-panel multiple giving a maximum reach of $25\frac{1}{2}$ inches to insure operating efficiency.

In determining the maximum vertical reach for the operator, the standard practice was followed of limiting this reach to 30 inches for line terminations which are to be answered, and to 34 inches for lines to which calls are to be completed. The line terminations to be answered by the operator are kept lower than the lines for completing purposes because the operator's attention must be attracted to the line by the illumination of the line lamp. The line capacity of the switchboard is limited by the number of line terminations that can be provided within the above dimensions.

The lower line of Fig. 3 shows the inter-toll trunk and subscriber line capacities obtainable within the permissible reach limits on a 6-panel multiple basis where the complete subscriber multiple is equipped with answering lamps. The capacity shown is based on various ratios of subscriber lines to inter-toll trunks. Because of the essentially toll character of the teletypewriter exchange service, it was anticipated that there would be a high ratio of inter-toll trunks to subscriber lines and comparatively little local traffic. The traffic studies indicated that the average ratio would be in the order of seven or eight subscriber lines

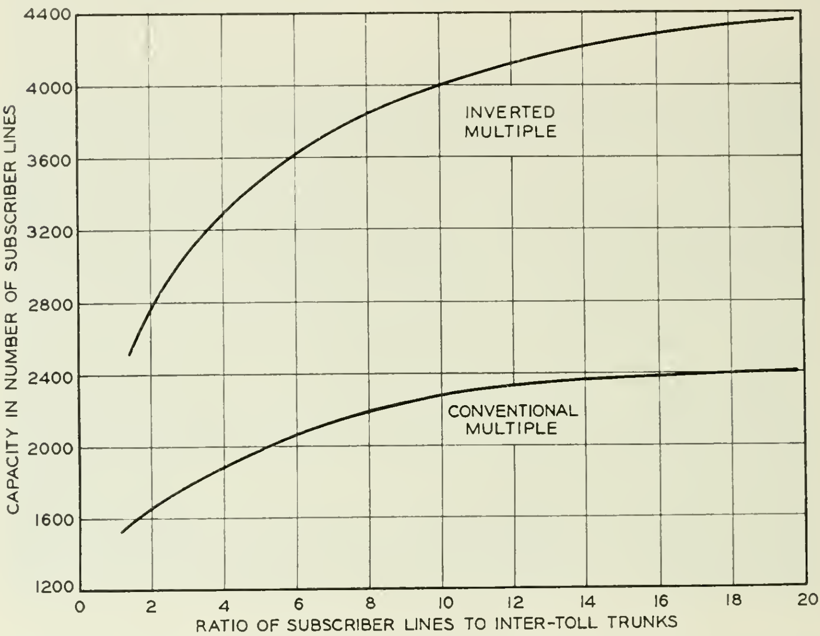


Fig. 3—Curves showing variation of subscriber line capacity for No. 1 Teletypewriter Switchboard. *A*—Inverted multiple. *B*—Conventional multiple.

to one trunk. It may be seen from Fig. 3 that, with this ratio, a capacity of only 2,200 subscriber lines is obtainable with the entire multiple equipped with answering lamps.

By providing answering lamps for only the first half of the subscriber lines and installing the second half without answering lamps in the upper portion, the total space for a given number of lines can be reduced and advantage taken of the additional space afforded by the 34-inch vertical reach permissible for calling multiple. This arrangement, known as the inverted multiple, provides answering facilities for

the second half of the subscriber lines in a second line-up of switchboard in which they are equipped with answering lamps. The first half of the lines also are multipled in this line-up but on a calling-only basis; that is, without lamps. With this arrangement, calls originated by the first half of the subscribers are answered in the first line of switchboard, and those originated by the second half are answered in the second line. Any operator may complete a connection to any line as there is a full multiple of jacks in both boards. In the second line-up the two halves of the multiple are inverted as to location in order to place the lines with answering lamps within easy reach. The upper line of Fig. 3 shows the capacities obtainable with this arrangement. It may be seen that, with a ratio of 7.5 subscriber lines to one inter-toll trunk, a subscriber line capacity of approximately 3,800 is possible. It was necessary, however, to reduce this to 3,600 lines in order to obtain a division in a multiple of 600 lines to simplify the numbering of the jacks. With this arrangement 300 lines are provided in each panel without answering lamps and 300 lines with answering lamps.

This multiple arrangement is illustrated in Fig. 4, which shows schematically the cabling for the first half of the subscriber multiple (lines 0 to 1,799). It may be noted that a third line of switchboard, the inward and through board, is provided. Experience has shown that the most efficient operation is obtained if the inward and through traffic is segregated when the switchboard grows to 30 or more positions. As the subscriber multiple is used here for calling purposes only, the answering lamps may be omitted from the entire subscriber multiple, thus making additional space available for increased inter-toll trunk capacity as discussed in the following. The subscriber lines are cabled from the main distributing frame (*MDF*) to the relay equipment and from there to the *TWX* intermediate distributing frame. Here cross connections are provided to permit the assignment of any subscriber line relay equipment to any multiple jack for flexibility in assigning numbers. The distributing frame terminal strip also serves as a doubling-up point for the cable to the switchboards.

A somewhat similar arrangement used for the inter-toll trunk multiple is shown schematically in Fig. 5. The standard telegraph line facilities and terminating repeaters designed for private line service are used for the *TWX* trunks. Connections to these trunks are made at the test board distributing frame and the trunks are cabled to the *TWX* distributing frame. Here arrangements are provided for inserting a single-line repeater, which is necessary for converting the positive and negative 130-volt signals to positive and negative 48-volt

signals for transmission through the switchboard. The trunk is then carried to the teletypewriter test board where a termination is provided for the purpose of testing the equipment. From the test board the trunk is cabled through the relay equipment to the distributing frame, where it is cross-connected to the switchboard multiple, the multiple for all three lines of switchboard being doubled up at the distributing

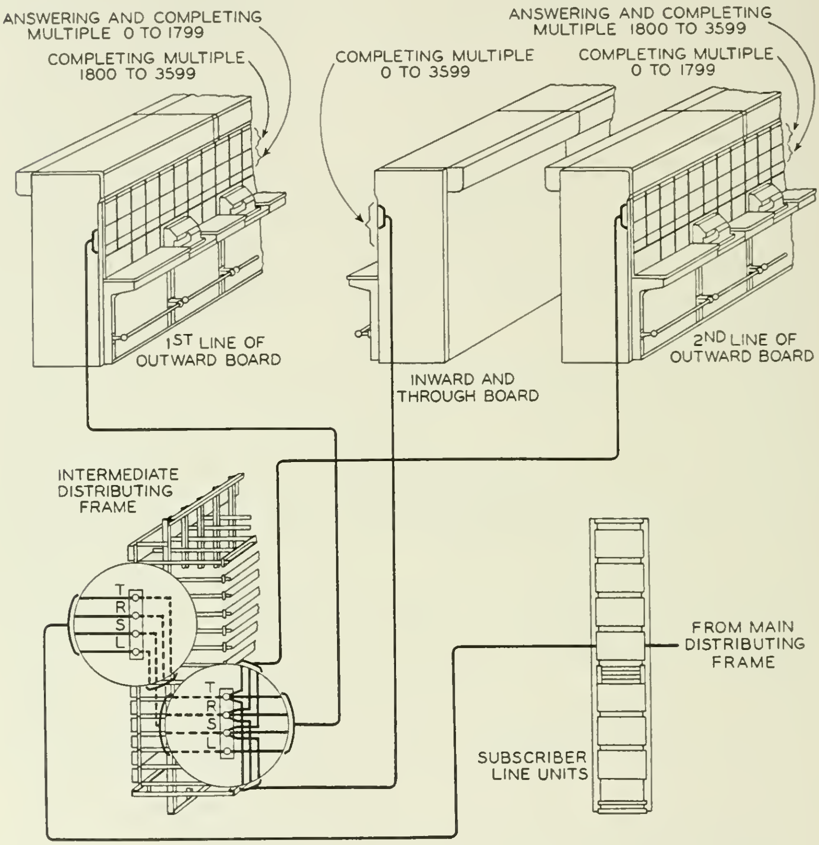


Fig. 4—Diagram of cabling for subscriber lines on No. 1 Teletypewriter Switchboard.

frame terminal strips. Ordinarily, with the inverted subscriber multiple arrangement, there will be a capacity for 480 inter-toll trunks equipped with answering lamps in the first two lines of switchboard. However, opportunity is provided for increasing this capacity by the provision of the separate inward and through switchboard. As described above, the lamps in this inward board may be omitted from

the subscribers' multiple. This arrangement provides sufficient space for the installation of 840 trunks equipped with answering lamps. As all trunks are answered at the separate inward switchboard, the answering lamps may be removed from the trunk multiple in the two

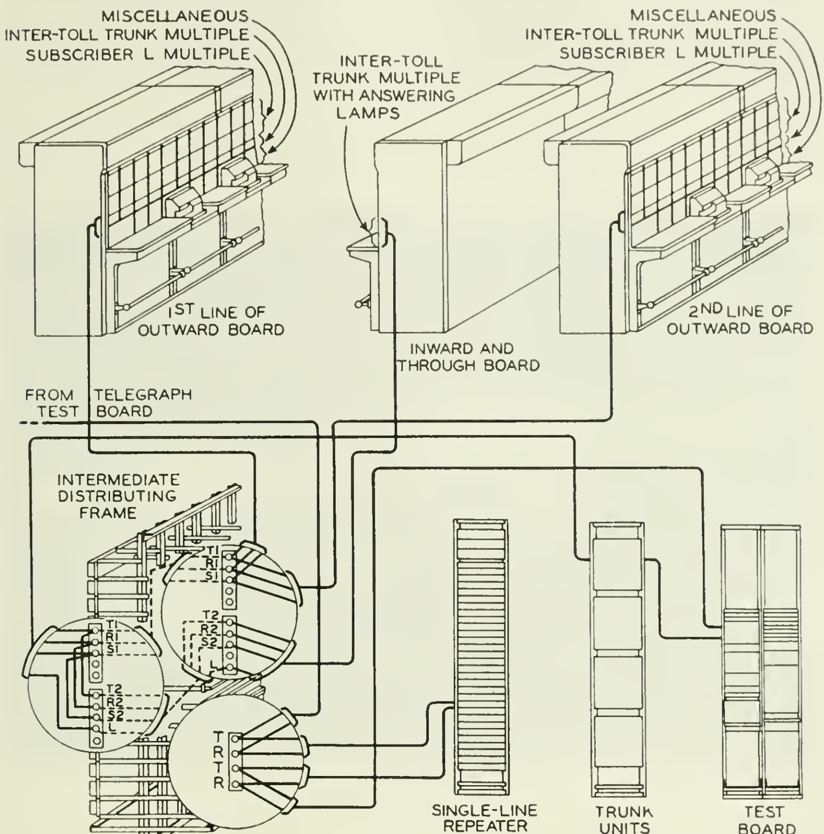


Fig. 5—Diagram of cabling for inter-toll trunks on No. 1 Teletypewriter Switchboard.

outward switchboards, thereby releasing sufficient space for the full 840 trunks without the answering lamps.

No. 3A Teletypewriter Switchboard

The design of a switchboard for the medium sized *TWX* switching points was not undertaken until the system had been in operation for about two years, temporary switching facilities having been used at these points in the meantime. Actual operating experience and

traffic data were then available upon which to base the design of the switchboard, a general view of which is shown in Fig. 6.

Efficient design requires that the width of a position be kept as small as possible to avoid the excessive cost of a long switchboard multiple. Because the smaller capacity required for this switchboard did not make the vertical reach an important factor, a key-shelf arrangement different from that used for the No. 1 switchboard was adopted.

Instead of placing the teletypewriter and cords on the same level as in the No. 1 switchboard, the cords are placed above the

Fig. 6—No. 3A Teletypewriter Switchboard at Pittsburgh, Pa.

teletypewriter. This was accomplished by the use of a sloping key-shelf permitting the cords to pass by the teletypewriter in the manner shown by the cross-sectional view in Fig. 7. With the object of keeping the vertical height of the keyshelf as small as possible, the cords are located in a single horizontal row instead of in the conventional double row. With this arrangement, the answering cord is the left cord of a pair and the calling cord is the right cord. Differentiation is obtained by using colored plug shells, black for the answering cord and red for the calling cord.

An additional feature resulting from this relation of the keyshelf to the teletypewriter is an arrangement whereby the position may be

adjusted to include various numbers of cords. This flexibility is obtained by the location of the teletypewriter on a table separate from the switchboard, connections being made by a flexible plug-ended cord. This permits the location of the teletypewriter in front of any group of cords. A position jack is provided in each section which affords facilities for operators spaced on minimum centers of $20\frac{1}{2}$ inches, each operator having access to a maximum of 10 cords. This repre-

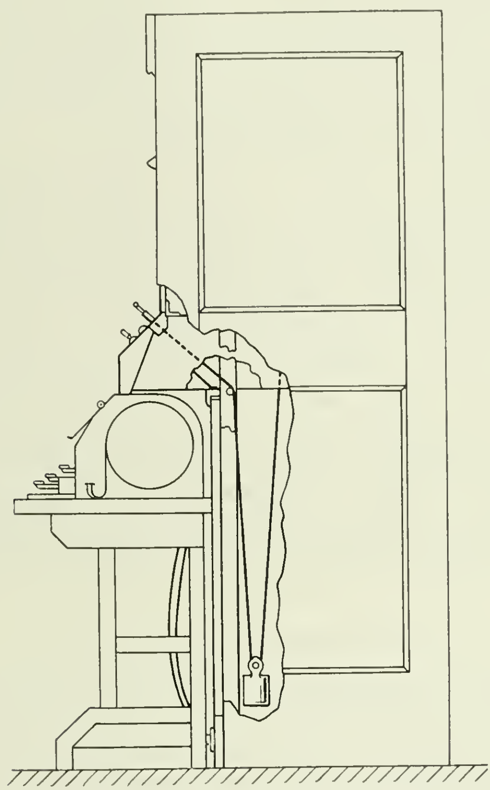


Fig. 7—Sectional view of No. 3A Teletypewriter Switchboard.

sents the closest centers which can be obtained with sufficient physical room for operating. Although the switchboards are usually engineered on the more ample operating centers of about $25\frac{5}{8}$ inches, the design permits the reduction of these centers to the $20\frac{1}{2}$ inch dimension in the event that more operators are required for unexpected increases in traffic. If traffic conditions change or the inward and through traffic is segregated, thus necessitating positions equipped with more

cords, the width of the position can be increased to include the number of cords required.

The switchboard is divided into sections, each having two panels and each arranged for a position circuit. The section is an arbitrary division of the switchboard for constructional purposes and has no bearing on the position boundaries. All keys and cords in a section are terminated on terminal strips in the rear. The cord relay equipment is furnished in units of 10 circuits, each unit being equipped with terminal strips so located that, when a unit is placed in the rear of a section, the terminal strips come directly under the section terminal strips. Distributing rings above the two rows of terminal strips provide facilities which permit any relay equipment to be cross-connected to any keyshelf cord equipment.

The engineering of this switchboard is thus reduced to a very simple process. The number of cords required per operator is determined by the anticipated traffic data. From this information the width of each position is determined. The sum of the positions required to handle the peak load represents the total length of the switchboard and determines the total number of sections required. Cord units are then provided in the rear of the switchboard. The cords required for each position are then cross-connected to relay circuits on the cord units which are in turn cross-connected to the nearest position circuit. Teletypewriters are moved in front of the various groups of cords and plugged into the jacks for their position circuits. Should conditions require a different assignment of cords, they may be recross-connected to meet the new requirements and the teletypewriters moved to new positions.

No. 3A Teletypewriter Multiple Equipment

For convenience, the operator's vertical reach for lines with answering lamps has been defined as 30 inches above the standard type of keyshelf. From the lower edge of the keyself, which prevents the operator from rising to reach farther, the permissible reach is 35 inches. Deducting the space required for the teletypewriter and keyshelf equipment, there remains $14\frac{1}{2}$ inches available for multiple below the 35 inch reach limit. About $2\frac{5}{8}$ inches of this space is required for unattended line terminations and miscellaneous multiple, leaving a space of $11\frac{7}{8}$ inches for the subscriber line multiple.

This space provides for the capacities shown in Fig. 8, which are in terms of ratios of subscribers lines to inter-toll trunks. This curve is based on the use of a 6-panel multiple which, with the $10\frac{1}{4}$ inch panel required for the type 49 jack used, results in a horizontal reach of

$30\frac{3}{4}$ inches. It may be seen that, with a ratio of 7.5 subscriber lines to one trunk, a capacity of about 1,300 lines is obtainable. Because the ratio of trunks to subscriber lines is somewhat greater on small switchboards than on the larger boards, due to the relative inefficiency of smaller trunk groups, the multiple is designed on the basis of 1,200 subscriber lines and 240 inter-toll trunks which gives a ratio of five subscriber lines to one trunk.

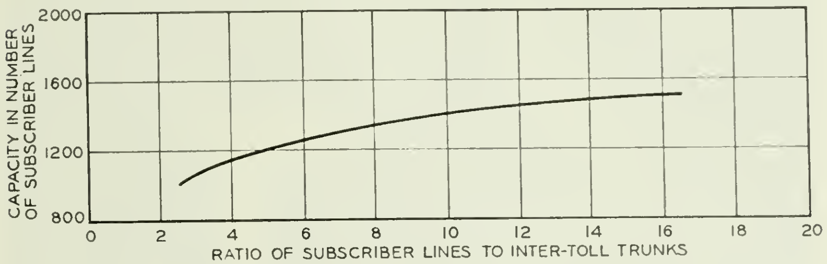


Fig. 8—Curve showing variation of subscriber line capacity for No. 3A Teletypewriter Switchboard.

Circuit Functions

The foregoing paragraphs have given a picture of the physical arrangement of the more important switchboards, and an idea of the number of subscriber lines and inter-toll trunks that can be accommodated by each. Some idea must also be given of the circuit methods by means of which connections are established between the various subscribers and supervised by the operators.

Subscriber Station and Station Circuit

The basic instrument by means of which the subscriber sends and receives his message is the teletypewriter. It is not proposed to give here a description of the teletypewriter as this is discussed in other papers. Other equipment, however, is required in addition to the teletypewriter to provide for the necessary signaling facilities for the exchange service.

A typical installation in a subscriber's office is shown in Fig. 9. The arrangement shown provides for the No. 15 (page) teletypewriter, used predominantly in the *TWX* service, mounted on a table which has been designed to provide adequate mounting facilities for the signaling equipment. This table is arranged with a removable panel known as a control panel, which is mounted in an opening in the top of the table to the right of the teletypewriter to make the key

levers readily accessible to the attendant. The control panel equipment may be varied to meet the different service requirements. Space is available on a shelf on the inside for a rectifier or an apparatus box where this additional equipment is necessary.

Fig. 9—Typical teletypewriter subscriber equipment for attended service.

A typical circuit arrangement for a station connected to a *TWX* switchboard is shown in Fig. 10. The station is equipped with a switch which, when operated, applies power to the motor of the teletypewriter, and also closes the loop so that a relay in the central office is energized. This relay lights the answering lamps associated with the subscriber's multiple in the face of the switchboard. An

operator answers by plugging the answering plug of a cord circuit into the jack. This action by the operator connects the station line to the cord circuit, and in addition energizes another winding of the relay previously energized when the subscriber called. This relay, being differentially wound, is then released and the answering lamps are extinguished.

In addition to calling the central office the subscriber must be able to recall the operator in case new services are required during the progress of the communication. This is accomplished by the subscriber simply turning the power switch off and then on again, which causes certain relays in the cord circuit to operate and the cord lamp to flash intermittently, indicating to the operator that her services

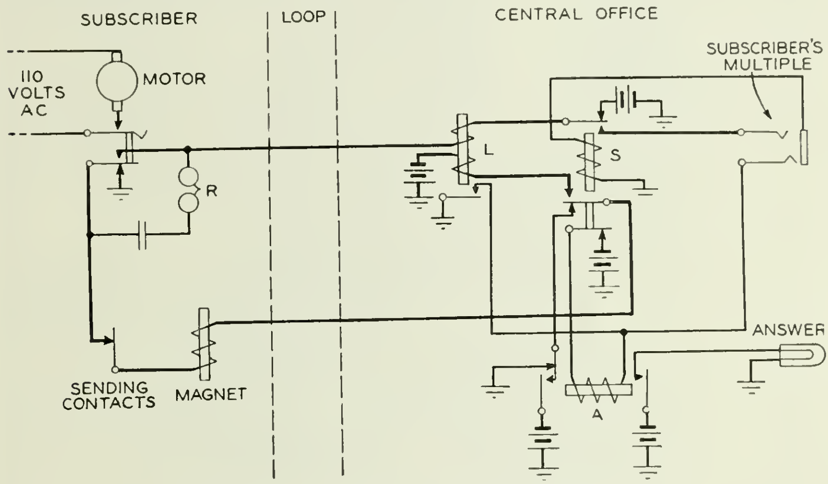


Fig. 10—Fundamental teletypewriter circuit.

are required. The subscriber must also be able to indicate to the operator when a disconnection has been made. This is accomplished by the subscriber turning off the power switch, causing the motor of the teletypewriter to stop and a lamp in the cord circuit to light.

The operator must also be able to signal the subscriber that a call is being completed to him. To provide this signal the station is equipped with a standard telephone type ringer which is energized, when the station is in the idle condition, by 20-cycle alternating current which flows over one side of the loop when the operator depresses a ringing key in the cord.

The teletypewriter lends itself admirably to the function of leaving messages on the subscriber's machine when no one is in attendance. When such service, known as unattended service, is desired, the station

is similar to that already described, but additional equipment is provided for starting the motor from the switchboard. An attempt is made to complete the call on an attended basis as outlined above and, if the called subscriber does not answer, the operator asks the calling subscriber if he wishes to leave his message. If he does, she presses a key in the cord circuit, which starts the motor at the absent subscriber's teletypewriter. The operator then instructs the calling subscriber to proceed with the communication.

Long Subscriber Lines

The subscriber stations just discussed are connected to the central office by two wires, known as a loop, the maximum distance between station and switchboard for loop connections being approximately 38 miles. A network is placed in the loops where the mileage makes its use necessary to improve the transmission efficiency.

It is necessary in some instances to connect subscribers situated at greater distances from the switchboard, perhaps as much as 200 or 250 miles. Two methods are available for accomplishing this: the d-c. method using telegraph facilities, and the carrier method using telephone facilities.

With the d-c. method a standard telegraph repeater is used at the central office, and a simplified repeater is placed on the subscriber's premises. These repeaters, with suitable signaling apparatus, provide a high grade of transmission and also the same type of supervisory signals as would obtain on the shorter loop connection.

The carrier method is used to a limited extent in the few instances where telegraph facilities are not available. In this method both the central office and the station are equipped with carrier apparatus and the regular telephone facilities are used. When the subscriber operates the power switch of the station, an answering lamp is lighted before the operator of the local telephone switchboard. The local operator, knowing by the multiple marking that this is a teletypewriter station, immediately connects through to the *TWX* switchboard over the regular toll telephone facilities. When the *TWX* switchboard is reached the call is handled in the same manner as a regular d-c. telegraph connection. All the signaling facilities available for the other subscriber stations are also available here. Completion to the subscriber is also made by the *TWX* operator over the regular toll telephone facilities, and the local telephone operator at the switchboard to which the subscriber is connected rings the subscriber.

Inter-toll Trunk Supervision

To provide inter-toll trunk supervision in the *TWX* network, it was necessary to select different types of signals than those occurring

during the normal transmission period of the teletypewriter; that is, the code and "break" signals.

Three types of supervisory signals are required to be sent over the inter-toll trunk. These are (1) the call signal, (2) the recall signal, and (3) the disconnect signal. There is a fundamental difference, however, between the call signal and the others in that it is applied to the trunk only when the stations are not connected. When the stations are connected the apparatus for receiving the call is removed from the trunk. The calling signal can therefore be any type of signal

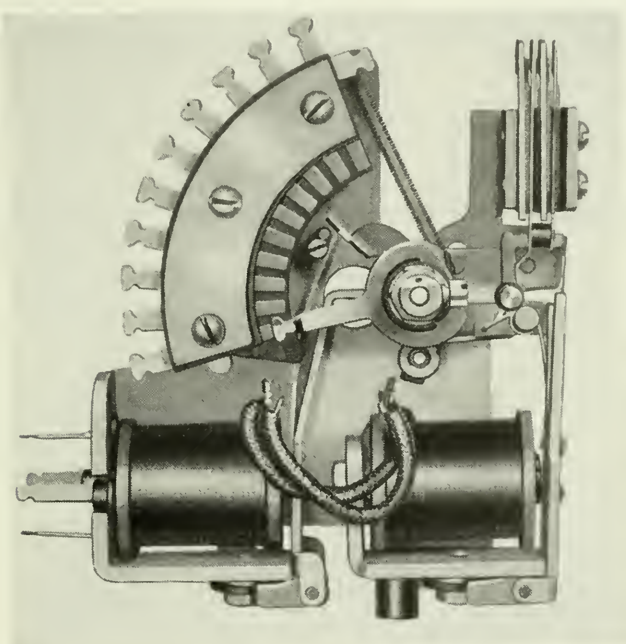


Fig. 11—Selector used for timing supervisory signals.

with the limitation that it must be such that it will not be produced by ordinary interruptions of the line, or line "hits."

The three types of supervisory signals chosen are therefore:

1. The call signal, produced by sending a spacing signal of 2 seconds.
2. The recall signal, produced by sending a spacing signal of 7 seconds.
3. The disconnect signal, produced by sending a spacing signal of 10 seconds.

To permit sending these signals, use is made of a mechanism that will measure the length of the signal. The basic apparatus used to measure this time is the selector shown in Fig. 11. By the use of this selector in conjunction with a ground interrupted 60 times per minute,

it is possible to obtain a means of measuring a line "open" within a sufficient degree of accuracy.

A typical method of sending and receiving the timed spacing signals or "opens" is shown in Fig. 12. The method of sending is shown at (A), and the method of receiving at (B). To send a recall signal, the operator at (A) presses the toll signal key momentarily leaving the cord up, the sleeve relay therefore remaining operated. The closure of the toll signal key operates relay A. Relay A operated opens the loop circuit at both ends of the trunk, releasing relays B and C. Relay B released closes a circuit which causes the selector to step at the rate of 60 steps a minute. The release of relay C causes relay D to release and provide a circuit for the selector at (B) to step at the same rate. When the selector at (A) reaches the first point it locks relay A and both selectors continue to step until the selector at (A) reaches the seventh point, when the locking circuit of relay A opens and that relay releases, closing the circuit and reoperating relays B, C, and D. The reoperation of relay B energizes the release magnet of the selector through the off normal contacts which cause the selector at (A) to release. At (B), when the selector reaches the sixth point, relay K operates and, when relay C reoperates, ground is connected through the contacts of relays K, L, and M to operate relay N. Relay N connects ground to the cord lamp, lighting it. Relay N when operated also connects ground to relay M which locks under control of contact P. When relay C reoperates ground is also connected to relay D which reoperates. Battery is then connected to the release magnet and the selector releases. After a time relay K, which is slow to release, also releases causing relay N to release. The release of N connects ground interrupted at the rate of 60 times per minute to the lamp which flashes until contact P is opened by the typing key, releasing relay M.

To send a disconnect signal, the same operations take place at (A) except that, immediately after the cord key is operated, the cord is pulled down, releasing the sleeve relay, and causing the selector at both ends to continue to the tenth point. At (B) when the selector reaches the tenth point relay L operates and, when relay C reoperates, ground is applied to operate relays M and N which hold a steady ground on the cord lamp until the cord is pulled down.

These signals appear at all offices in a built-up connection. The frequency of the machines supplying the 60 interruptions per minute is accurate to within plus or minus five per cent, and the multiple connections on the receiving selector bank take up any inequalities that may exist in the speed of the machines in two different offices.

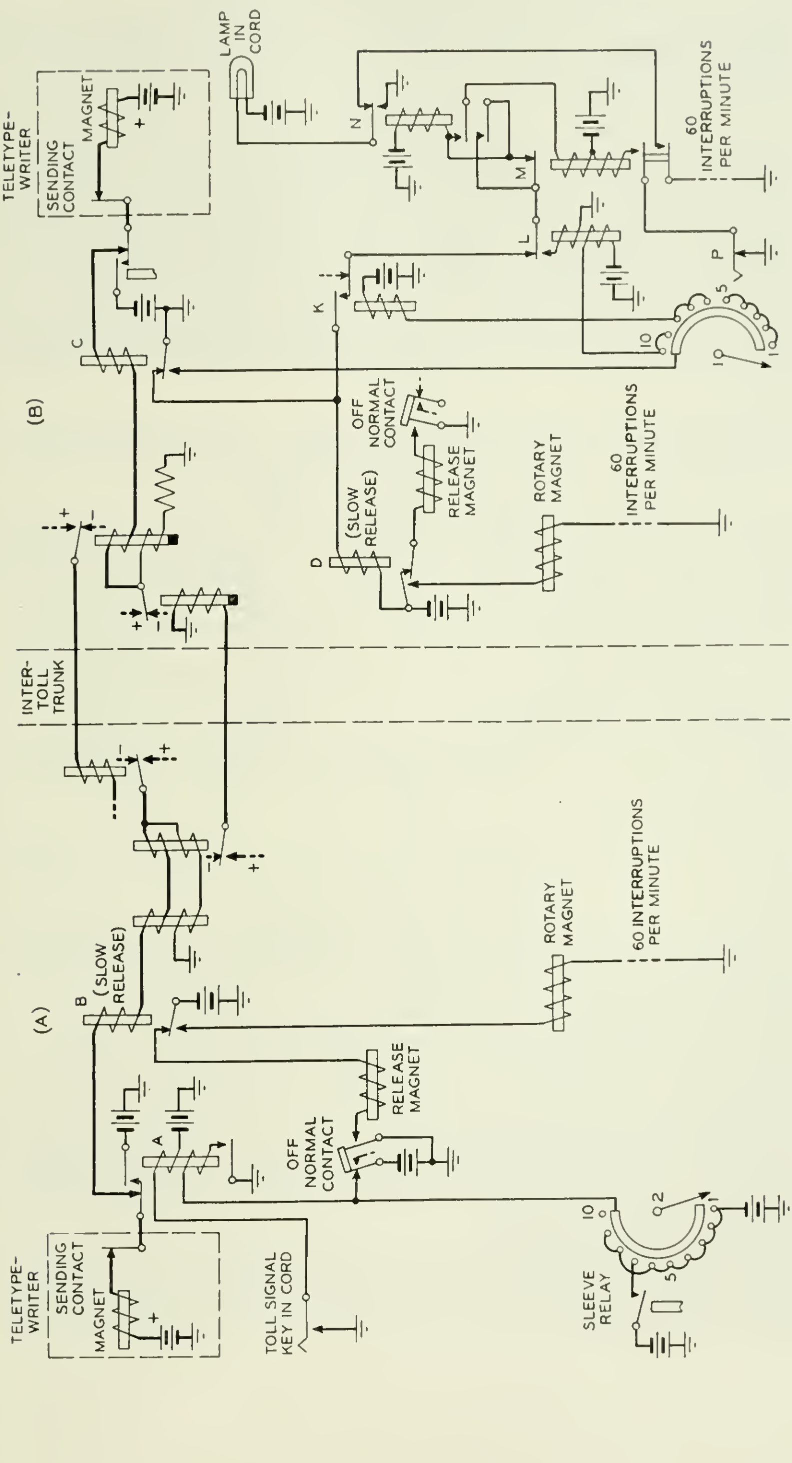


Fig. 12—Fundamental inter-toll trunk signaling circuit.

That section of the receiving selector between terminals 1 to 5 inclusive is used for the call signal which is actuated manually by the originating operator.

Cord Circuits

In order to provide each operator with sufficient traffic for operating efficiency, especially in the smaller offices and during light load periods, the cord circuits in the *TWX* switchboards are made universal, that is, adapted to handling all types of calls. This universal feature is obtained by equipping them with a simple type of repeater. By means of this repeater it is possible to provide for the maximum length of station loop and at the same time establish the following connections:

1. Subscriber line to subscriber line, known as a local to local connection.
2. Inter-toll trunk to subscriber line, or *vice versa*, known as a toll to local connection, or local to toll connection.
3. Inter-toll trunk to inter-toll trunk, known as a through connection.

In a local-to-local connection the two loops could not be connected together directly unless the repeater were provided in the cord circuit for two reasons: first, each loop may be maximum in length so that the two loops in tandem would result in the operating current being halved; and second, each loop is normally terminated on the negative side of the telegraph battery. Because it is essential, in *TWX* service, to make interconnections without requiring adjustments, all loops are padded or "built out" to the same value as the resistance of a maximum loop and each side of the cord circuit repeater is arranged to operate with each loop.

With the provision of the repeater in the cord circuit to permit interconnecting two subscriber lines, the same cord may be used for toll-to-local and toll-to-toll connections because the loop circuits of the inter-toll trunk repeaters are all terminated on the negative side of the telegraph battery and the loop resistance of each is built out to equal that of the longest station loop.

A very simplified form of the essential elements of a *TWX* cord circuit is shown in Fig. 13. The cord circuit basically consists of a repeater of the type before mentioned, a key known as the typing key, by means of which the operator may cut her teletypewriter in and out of the circuit for monitoring purposes, and facilities for receiving the recall and disconnect signals both from the subscriber lines and the inter-toll trunks.

The method of receiving these recall and disconnect signals was explained in the section on inter-toll trunk supervision, and the method used to receive those from the subscriber was pointed out in the

subscriber circuit description. Many other items are included in the cord circuit by means of which the operator may expedite the setting up and removing of connections. Among these items is the busy test. When an operator is about to complete a call to a station it is necessary that she know that the station is free to receive the call. To ascertain this a means is provided so that she may make a tip busy test on the sleeve of the jack associated with that subscriber line and, if the station is busy, a position light will be lit. If no light is received the operator will plug into the jack and complete the connection.

Multiple appearances of the jacks and lamps associated with subscriber lines and inter-toll trunks are provided so that a number of

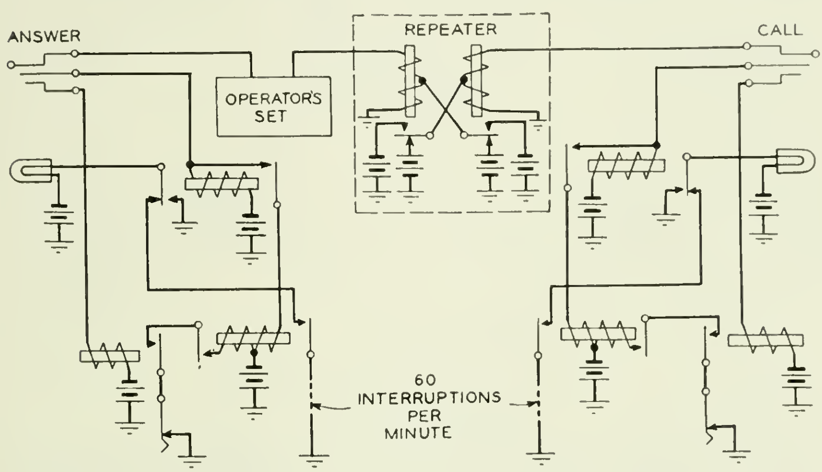


Fig. 13—Typical cord circuit.

operators may be available to answer a call from a station or an inter-toll trunk. If more than one operator answers it is necessary that they be aware of that fact, and that the first operator shall take and complete the call. A circuit is provided to indicate this.

Facilities are provided to split the cord, that is, to enable the operator to communicate in one direction without the communication being recorded in the other direction. Ringing is accomplished in a manner similar to that used in telephone practice, the No. 1 switchboard using manual start machine ringing and the smaller No. 3*A* switchboard using manual ringing. While the cord is connected to one line and the operator is attempting to complete the connection to another line, the first line is held closed in order not to mar transmission.

Conference Connections

The teletypewriter exchange system provides a means whereby practically unlimited numbers of stations can be connected in conference connections. Figure 14 shows a typical conference connection. Each link in the conference circuit is provided with a simple repeater, each of which is equipped for breaking. In this manner to and fro communication by the half-duplex method operation can be attained. The conference repeater circuits are made up in groups of five or ten, each of which is equipped with a multiple appearance so that all operators have access to the repeaters.

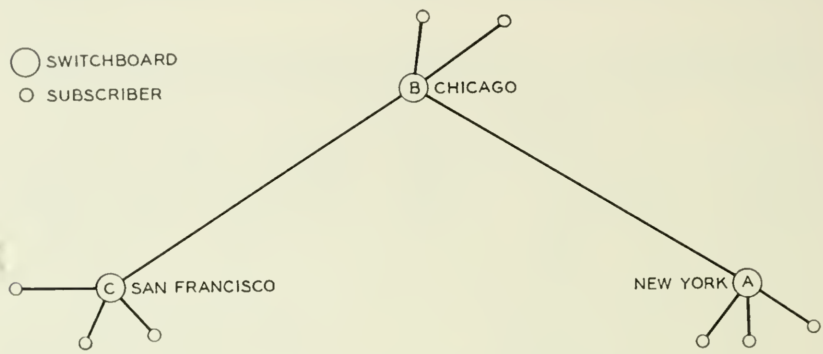


Fig. 14—Typical conference connection.

Regenerative Repeaters

It is necessary in some long circuits to improve transmission by inserting a regenerative repeater in the circuit. To make this possible and easily performed by any operator, regenerative repeaters are provided with a complete jack multiple appearing before all operators. Both ends of the repeater are available in this multiple and the repeater may be inserted where the transmission equivalent of the circuit involved makes it necessary.

Typical Built-Up Connection

In order to provide *TWX* service on a nationwide basis, certain of the connections require one or more intermediate switchboards so that one or more through operators may be involved. As an illustration of this Fig. 15 shows a connection established between a calling station in New York and a called station in San Francisco with a through switch at Chicago, a method used when all direct trunks are busy. This figure shows the manner in which the *TWX* equipment

has been arranged to operate in conjunction with the telegraph line facilities. The station loop is a pair of wires such as those used for telephone service. Each inter-toll trunk consists of one or more sections of the same standard types of carrier, metallic, or grounded telegraph line systems that are employed in private line telegraph service. The signaling and supervisory apparatus is all contained in the *TWX* switchboard equipments.

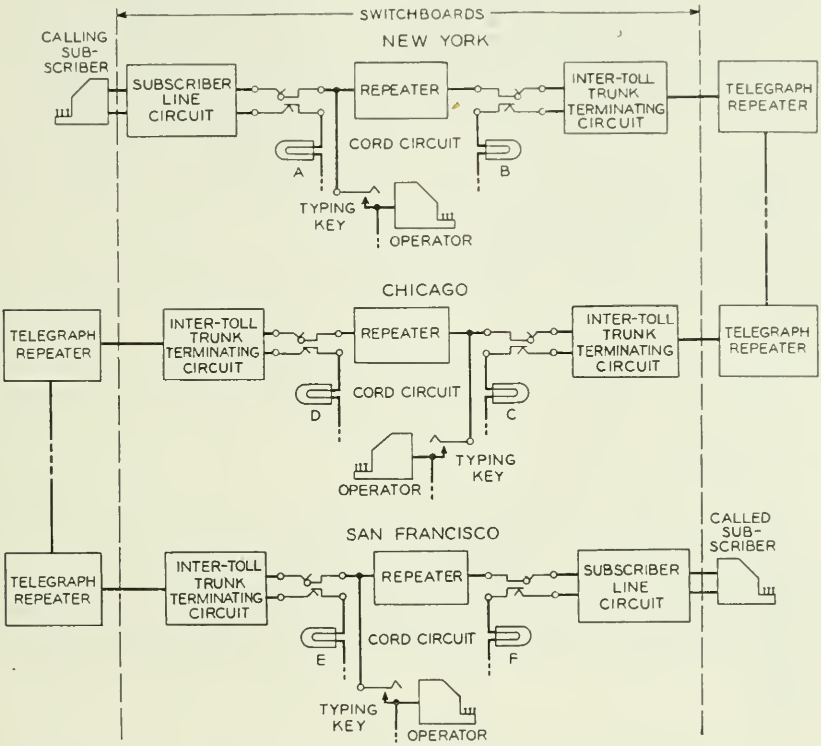


Fig. 15—A built-up connection such as would be used if all direct New York-San Francisco trunks were busy.

In the example illustrated the New York operator, being the outward operator, supervises the call and times the ticket. The following traffic table shows the important steps taken in setting up and taking down this connection:

1. *The New York subscriber calls:* Subscriber closes loop and starts teletypewriter by operating switch. Subscriber line lamps light in the New York switchboard.

2. *A New York operator answers with the cord typing key (similar in function to the talking key in the telephone cord circuit) operated:* The line lamps are extinguished. The operator and subscriber communicate.

3. *The New York operator connects to an idle trunk in the New York-Chicago multiple:* Plugs the completing end of the cord into the trunk jack and operates the cord ringing key for 2 seconds. The trunk multiple lamps at Chicago light.

4. *A Chicago operator answers:* Plugs the answering end of a cord into the trunk jack with the cord typing key operated. The trunk lamps are extinguished. The Chicago operator communicates with the New York operator.

5. *The Chicago operator completes to an idle trunk in the Chicago-San Francisco multiple:* Plugs the completing end of the cord into an idle trunk jack and operates the cord ringing key for 2 seconds. Releases typing key. The trunk multiple lamps at San Francisco light.

6. *A San Francisco operator answers:* Plugs the answering end of a cord into an idle trunk jack with the cord typing key operated. The trunk lamps are extinguished. The San Francisco operator communicates with the New York operator.

7. *The San Francisco operator completes the connection to the called subscriber line:* After making a tip busy test with completing cord to insure that the called station is idle the San Francisco operator plugs into the jack and operates the ringing key in that cord. The ringer in the San Francisco station is operated.

8. *The called subscriber answers:* The answer is received on the operators' teletypewriters at the San Francisco and New York switchboards and at the New York subscriber station. The San Francisco and New York operators release the cord typing keys leaving the communication between the subscribers. The New York operator starts timing the ticket.

9. *The calling and called subscribers disconnect:* Lamps A and F light. The New York operator completes the timing of the ticket.

10. *The outward (New York) operator sends the disconnect signal:* Operates cord key momentarily and pulls down both cords. After 10 seconds lamps C, D, and E light.

11. *The inward (San Francisco) and through (Chicago) operators disconnect:* Upon noting the disconnect lamp signals both operators pull down both cords.

If during the progress of the call the subscriber desires to regain the attention of the operator, a recall signal is sent. The procedure in this case is as follows:

12. *The calling (or called) subscriber recalls:* Operates power switch at the station. Cord lamp A (or F) flashes.

13. *The operator answers the recall:* Operates cord typing key connecting her teletypewriter to the circuit. The flashing cord lamp is extinguished.

14. *The outward (New York) operator recalls the inward and through operator at Chicago:* Operates recall key in cord. After 7 seconds lamps B, C, D, and E flash. The outward operator releases the typing key momentarily to extinguish the lamp.

15. *The inward and through (San Francisco and Chicago) operators challenge:* Operate cord typing keys which extinguish the flashing lamps, and then challenge by typing.

Conclusion

This paper has outlined the technique of teletypewriter switchboard operation as it stands today. Although the designs as here outlined have given satisfactory service within due limits of economy, the expansion of the system and experience in its operation will undoubtedly lead to changes in the design of both the equipment and circuits and to changes in the methods of operation to increase the efficiency and improve the quality of the service.

A Transmission System for Teletypewriter Exchange Service *

By R. E. PIERCE and E. W. BEMIS

A nationwide transmission system has been established in the United States for teletypewriter exchange service by means of which 2-way communication between teletypewriter subscribers can be established in a time comparable to that required for long distance telephone service. A brief description of the principle of operation of teletypewriters is included in this paper as an introduction to the discussion of the transmission requirements and the plan of the present system.

TO MEET the growing needs of business organizations, particularly those operating on a nationwide basis with branches at widely separated locations, there has developed in the United States an extensive use of private line telegraph service. This trend has been accelerated by the perfection of the teletypewriter, which makes it possible for regular office employees to transmit and receive communications without a large amount of special training. Some of these private line teletypewriter networks have been provided with switching facilities to permit the customer to set up connections between his various offices or groups of offices as desired. As these arrangements were perfected and as the public gained experience with the teletypewriter method of communication, a demand developed for a teletypewriter service in which all connections would be set up on a switched basis similar to that provided for spoken conversation by the telephone system. To meet this demand teletypewriter exchange service or as it is usually called, *TWX* service, was inaugurated by the Bell System in November 1931.

Briefly described, teletypewriter exchange service makes available to subscribers a complete communication system for the written word, consisting of:

(*a*) Teletypewriters for sending and receiving, installed on the customers' premises with a connection to a nearby switching center.
(*b*) Transmission channels interconnecting all of the switching centers.
(*c*) Teletypewriter switchboards for connecting the subscribers' stations and loops to each other or to the inter-city transmission channels and for making through connections between inter-city circuits.

This system provides for direct teletypewriter connections between the customers or their employees at the sending point and at the receiving points. The connection is two-way so that questions can

* Published in *Electrical Engineering*, September, 1936. Presented at A.I.E.E. Southwest District meeting, Dallas, Texas, October 26–28, 1936.

be asked and answers given. The speed with which the connection is established is comparable to that experienced in long distance telephone service, the average being about 1.3 minutes from the time a subscriber calls the operator until the conversation between subscribers begins. The service has grown until at the present time there are over 8,500 * subscriber stations which may be connected together in pairs or in groups for teletypewriter communication. The switching is done at about 150 switching centers scattered throughout the United States as shown in Fig. 1 and connected by over 500,000 miles of telegraph circuit.

This paper deals primarily with the transmission system used for passing the teletypewriter code signals between the customers. The details of the switchboards and signaling facilities, and the methods of handling customers' connections are described in another paper.[1] With the exception of the switchboards, the equipment used in *TWX* service is similar to that used in other telegraph services.

The teletypewriters are provided with a keyboard similar to that of a typewriter for sending, and the typing is done in capital letters either on a narrow tape or on a page, the page being used in the large majority of the stations. Printed forms may be used on the page machines if desired. The speed of operation is set for a maximum of 60 words per minute. The teletypewriters are of the start-stop type, using a 5-unit selecting code, each group of selecting impulses being preceded by a start impulse and followed by a stop impulse. The teletypewriter mechanism is operated from a local source of power, and in general all signaling current is furnished from central office power plants.

The line circuits may be any of the well known types utilizing 2 current values or line conditions of variable duration for the transmission of signals. Actually about 90 per cent of the circuit mileage used in the *TWX* service is of the carrier type, since this is the most economical type of facility for large groups over the longer distances. The line circuits will be discussed in more detail in another section of the paper.

Elements of Teletypewriter Signal Transmitting and Receiving Mechanism

To translate intelligence which is received in the form of a code the receiving mechanism must be capable of doing two things. First, it must identify the unit time intervals, and second, it must determine which of the 2 line conditions should be recorded for each time interval.

* Since this paper was prepared the number of subscriber stations has increased to over 9,500.

[1] For all numbered references see list at end of paper.

be asked and answers given. The speed with which the connection is established is comparable to that experienced in long distance telephone service, the average being about 1.3 minutes from the time a subscriber calls the operator until the conversation between subscribers begins. The service has grown until at the present time there are over 8,500 * subscriber stations which may be connected together in pairs or in groups for teletypewriter communication. The switching is done at about 150 switching centers scattered throughout the United States as shown in Fig. 1 and connected by over 500,000 miles of telegraph circuit.

This paper deals primarily with the transmission system used for passing the teletypewriter code signals between the customers. The details of the switchboards and signaling facilities, and the methods of handling customers' connections are described in another paper.[1] With the exception of the switchboards, the equipment used in *TWX* service is similar to that used in other telegraph services.

The teletypewriters are provided with a keyboard similar to that of a typewriter for sending, and the typing is done in capital letters either on a narrow tape or on a page, the page being used in the large majority of the stations. Printed forms may be used on the page machines if desired. The speed of operation is set for a maximum of 60 words per minute. The teletypewriters are of the start-stop type, using a 5-unit selecting code, each group of selecting impulses being preceded by a start impulse and followed by a stop impulse. The teletypewriter mechanism is operated from a local source of power, and in general all signaling current is furnished from central office power plants.

The line circuits may be any of the well known types utilizing 2 current values or line conditions of variable duration for the transmission of signals. Actually about 90 per cent of the circuit mileage used in the *TWX* service is of the carrier type, since this is the most economical type of facility for large groups over the longer distances. The line circuits will be discussed in more detail in another section of the paper.

Elements of Teletypewriter Signal Transmitting and Receiving Mechanism

To translate intelligence which is received in the form of a code the receiving mechanism must be capable of doing two things. First, it must identify the unit time intervals, and second, it must determine which of the 2 line conditions should be recorded for each time interval.

* Since this paper was prepared the number of subscriber stations has increased to over 9,500.

[1] For all numbered references see list at end of paper.

Fig. 1—Map showing teletypewriter exchange service switching centers in the United States on April 1, 1936.

The first requisite is accomplished by maintaining a high degree of synchronism between the sending and receiving devices during the transmission of each character. The second is accomplished by providing satisfactory transmission facilities so that the mid-portion of each received signal element is the same as the corresponding signal element at the sending end.

Timing Arrangements

The sending and receiving devices are driven by motors which run at approximately the same speed. The receiving device is driven through a friction clutch so that it normally may be idle even though its motor is running. When a signal is received the receiving selector is released, makes one complete revolution, and again comes to rest. With this arrangement it is necessary to maintain synchronism only while one character is being transmitted, because a fresh start is made for each character, and the time intervals for the selecting impulses are measured from this starting point. Cumulative lack of synchronism, therefore, over long periods of time does not affect the accuracy of transmission. This is called the start-stop system.

The advantages of this arrangement are as follows:

(*a*) No elaborate means of synchronizing are required.
(*b*) The lag in the line is automatically taken care of because the receiving machine does not start until the first signal of a code combination is received.
(*c*) Multisection circuits and conference connections can be set up without any special line-up.
(*d*) Machines can be started and shut down without any special adjustment.
(*e*) Local power sources can be used for driving the subscriber's machine.

In actual practice speed is maintained within $\pm$ 0.75 per cent in either of two ways:

1. Where regulated frequency a-c. power is available, synchronous motors ordinarily will maintain the speed within $\pm$0.17 per cent, which is well within the limit necessary for satisfactory transmission.
2. Where regulated frequency a-c. power is not available, governed motors are used for either alternating or direct current. These governors are designed so as to maintain the speed within $\pm$0.75 per cent without attention over long periods of time. If the speed of the sending machine is out in one direction and that of the receiving machine in the opposite, the maximum difference may be 1.5 per cent.

Assuming no deformation of the wave shape between the sender and the receiver, the start-stop teletypewriter operating at 60 words per minute will stand about 7 per cent speed discrepancy before errors occur. In practice, however, there is deformation and therefore the speed discrepancy must be kept as low as practicable.

Sending and Receiving Arrangements

The sending arrangement in a teletypewriter is required to do three things:

1. It must transmit a signal which will start the selecting cycle of the distant machine.
2. It must apply the proper current condition to the line for each of the 5 accurately spaced selecting time intervals.
3. It must send a signal which will return the line to the normal idle condition.

The teletypewriter operates in a local circuit in which current is flowing during the normal idle condition. The transmitting is done by opening and closing this circuit, causing zero current or normal current in it, the two conditions being referred to as "open" and "closed." The selecting cycle of the distant machine is initiated by opening the circuit at the sending teletypewriter. This is called the "start" signal. The five selecting signals follow and the line current during each of these time intervals depends upon the character which is being transmitted. Since the normal idle condition of the line is closed, the "stop" signal which is sent last in the train of signals is a "closed" signal.

The selecting arrangement in a receiving teletypewriter is also required to do three things:

1. It must start timing the signals when the start signal is received.
2. It must determine the line condition at the midpoint of each selecting interval.
3. It must come to rest during the stop interval following the 5 selecting signals.

A single electromagnet in the receiving machine converts the electrical pulses into mechanical operations of the selecting mechanism. This magnet controls an armature which is energized for the closed line condition and de-energized during the open line condition. By this means the 2 line conditions are converted into 2 positions of the magnet armature.

Theory of Teletypewriter Signal Transmission

In teletypewriter signal transmission at 60 words per minute (hereafter called 60-speed) the start pulse and each of the 5 selecting signal elements are normally of 0.022 second duration. The minimum length of the stop pulse is 0.031 second. In keyboard sending the maximum length of stop pulse depends upon the time the operator hesitates between the striking of the individual keys of the teletypewriter. Any lengthening or shortening of the signal elements in transmission is referred to as distortion and is expressed as a percentage of the normal length of a signal element. The fundamentals of signal transmission have been discussed thoroughly by various writers.[2, 3, 4] A few of these principles are enumerated here without any attempt to discuss them thoroughly.

1. With the transmitting arrangements usually employed the complete change in line condition at the sender is practically instantaneous.

2. To transmit accurately these sudden changes in the line condition would require a transmission channel capable of passing an infinitely wide frequency band.
3. With a transmission channel which will pass only a limited band of frequencies there will be alteration of the wave shape during transmission as the result of changes in magnitude and phase of the various components caused by the characteristics of the transmission channel, so that changes in line condition at the receiving end will be gradual and in general displaced from their proper position.
4. Theoretically all of the intelligence can be carried by transmitting waves of a maximum frequency equal to that of the fundamental of the signaling speed considering the time interval of each signal element as a half cycle.
5. Actually it is not economical either to transmit a very wide band of frequency or to provide terminal apparatus capable of accurately recording the intelligence when only a band equal in width to the frequency of the fundamental of the signaling speed is transmitted. The arrangement used in practice must, therefore, be a compromise between these two extremes.

Experience has shown that in order to use economically practical types of receiving apparatus it is generally necessary to have present in the received signals a substantial portion of the second and third harmonics of the frequency of the shortest signal element, which requires in the case of 60-speed teletypewriter signals the transmission of a frequency band width of somewhat more than 45 cycles. To illustrate this a typical 60-speed teletypewriter signal is shown graphically in the upper left-hand diagram of Fig. 2. This diagram represents potential applied to the line for a perfect teletypewriter letter "D." At the instant when the start pulse commences, as described previously, the voltage applied to the line assumes its "open" value S, called "spacing." This spacing condition continues for 0.022 second at the end of which time the voltage suddenly assumes its "closed" value M, called "marking." The marking voltage remains constant through the first signaling pulse (1) in the figure. The second and third elements of the teletypewriter "D" signal are spacing and during these intervals the current is again of its spacing value. In the fourth pulse it once more becomes marking for 0.022 second, and in the fifth pulse it is again spacing. After the fifth pulse the current assumes its marking value for the duration of the stop signal.

This teletypewriter "D" signal may be further analyzed by considering it to be made up of sine wave components of various frequencies and magnitudes with certain definite phase relationships. It will be found theoretically to contain a number of sine waves of frequencies from zero to infinity. The left-hand column of Fig. 2 shows a number of the more important harmonic components of the "D" signal, the relative magnitudes and phase relationships being as indicated. The first is the d-c. component; the second is a sine wave of the same period as the over-all signal, and is referred to as the first harmonic. The wave shown in part *c* of the figure is twice the frequency of the over-all signal and is referred to as the second harmonic. Following this in turn are shown the third to tenth harmonics.

The right-hand portion of the same figure shows the synthesis of this signal from component parts. From this figure it may be seen that by the time the seventh harmonic (curve q) or even the fifth harmonic (curve p) has been added, there is a resemblance between the resultant and the original wave.

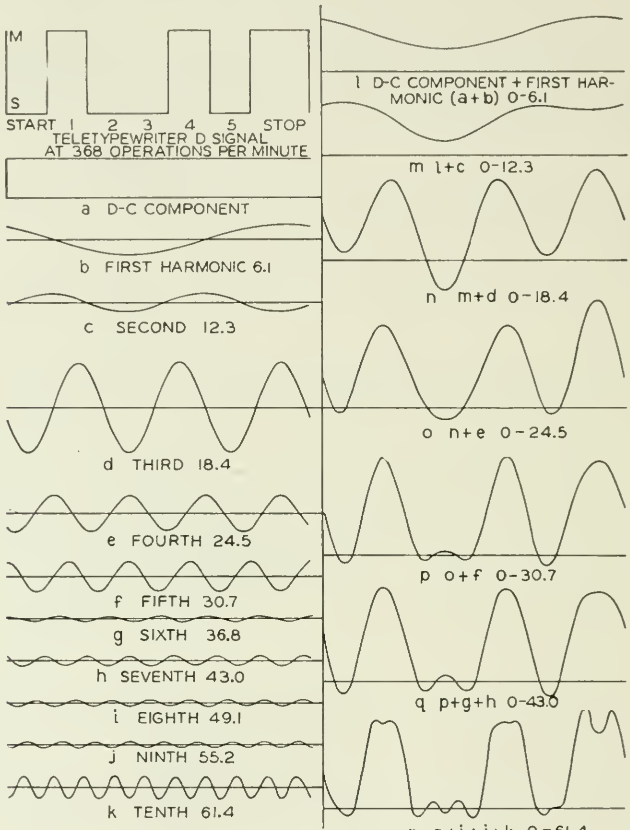


Fig. 2—Analysis of components of a teletypewriter "D" signal. Numbers on harmonic curves are frequencies in cycles per second.

As mentioned previously, the total intelligence transmitted by a given telegraph signal may be contained in a frequency band lying between zero and the fundamental frequency of the shortest signal element, i.e., the frequency at which the duration of the shortest element is a half cycle. In 60-speed teletypewriter transmission the shortest signal element is of approximately 0.022 second duration, and its fundamental frequency is about $1/0.044 = 22.7$ cycles per

second. In the illustration in Fig. 2, this frequency would fall between that of curves d and e in the left-hand column and the character theoretically could be interpreted correctly with the transmission in correct phase relation of only the components up to and including the fourth harmonic (curve o in the right-hand column). As previously stated, however, while transmission of such a limited frequency range could be interpreted without error by an ideal receiving device, practical considerations of over-all economy make it desirable to transmit the wider frequency range mentioned.

Types of Distortion

In order to design a satisfactory teletypewriter transmission system it is desirable to understand the effects of various types of distortion and mechanical variations in the sending and receiving mechanisms. Figure 3 shows schematically that part of the receiving mechanism which is of interest in explaining the effect of signal distortion on correct interpretation of the message. This includes a receiving selector magnet with its associated armature and armature extension, a locking lever, a stop latch, and a selector cam driven by a friction clutch. In the idle condition the selector magnet is energized and the magnet armature and armature extension are in the position shown, the selector cam being held from rotating by the stop latch. When a train of impulses representing a character is received the start pulse (spacing) allows the armature and armature extension to move to a position shown by the dotted lines and at the same time releases the stop latch. This latter operation permits the selector cam to start rotating. The speed of rotation and the starting position of the selector cam are normally so adjusted that the first depression (shown by A) will arrive at the locking lever at the time the middle of the first selecting impulse is being received. The locking lever will then fall into this depression and the locking wedge B will move toward the armature extension and lock it in the position it occupies at this instant. This determines which of the 2 line conditions will be recorded for this signal element. Immediately thereafter mechanical arrangements (not shown) will operate to transfer this information to the selection storing mechanism. This process will then be repeated for each of the other 4 selecting impulses.

After the 5 selecting impulses have been received the slightly longer stop impulse is received. During the latter part of this impulse an arm C on the receiving selector cam will strike the stop latch and the cam will be held until the reception of the start impulse for the next character.

An orientation device or range finder is provided which rotates the stop latch with respect to the locking lever and thereby changes the time at which selection occurs with respect to the beginning of the selecting cycle. Moving the orientation range finder in effect moves the solid vertical lines in Fig. 3, with respect to the signal, and with perfect signals they can be moved by an amount corresponding to one unit impulse. In other words the time of selection can be moved by $\pm$ 50 per cent without typing errors, as shown at *a* in the figure. (In an actual machine this range is less because of practical con-

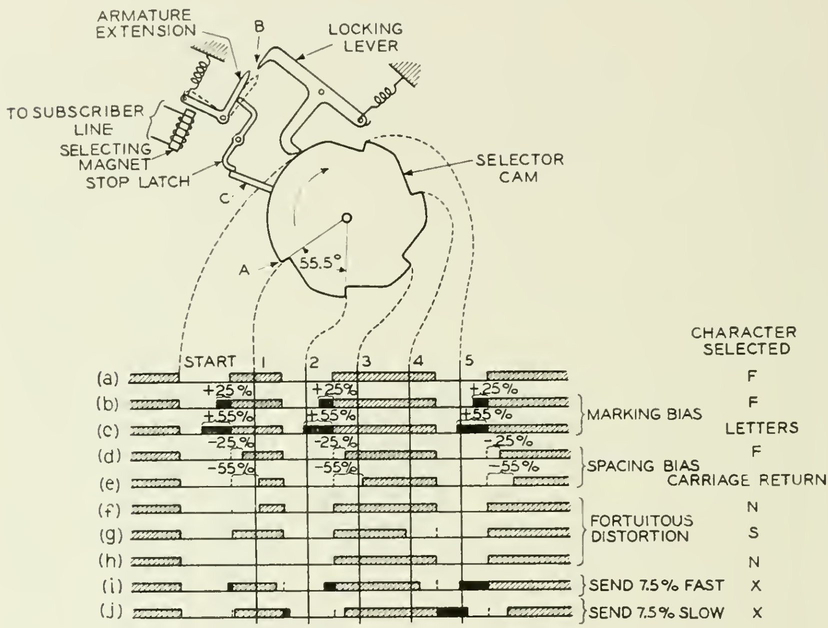


Fig. 3—Principles of selecting mechanism of a teletypewriter.

siderations of design, the time of selection being variable without errors over a range of about $\pm$ 40 per cent.)

Distortion in teletypewriter signals may be "bias," which is a uniform lengthening or shortening of all the marking impulses, or it may be of other types which affect only certain of the signal elements.[4] Bias is divided about equally between the ends of the impulse when the signal is received from the line. However, because the selecting mechanism starts rotating at the beginning of the start impulse, the effect of bias is to shift all impulses forward or backward with respect to this time. The result of this is that effectively there will be an

addition to or subtraction from the front of each marking impulse, with the rear of the impulses remaining unchanged.

In an ideal machine where selection would be made instantaneously the signal would be recorded correctly if it had the right condition (i.e., marking if it should be marking or *vice versa*) at the instant of selection. The particular times when the selections take place with the orientation setting at the middle of the range with perfect signals are shown, as mentioned before, by the vertical solid lines numbered 1 to 5, inclusive, in Fig. 3. Referring to cases *b* and *d* it may be seen that with 25 per cent bias the correct signal is being received at the point of selection and it will be interpreted correctly. However, referring to cases *c* and *e* it may be seen that more than 50 per cent bias will cause errors. In case *c* the second and fifth impulses will be falsely interpreted as marking and in case *e* the first and third impulses will be spacing instead of marking. Several examples of the effect of distortion other than bias in the received signals are illustrated in cases, *f*, *g*, and *h* of Fig. 3.

The effect of variations in teletypewriter motor speeds on operating margins is illustrated in Fig. 3 by cases *i* and *j*. Case *i* shows the result if the sending machine is faster than the receiving machine. It will be noted that as the speed discrepancy becomes greater the first error will be a false mark for the fifth impulse because a part of the stop impulse is received on the fifth position. If perfect signals are assumed, the speeds would have to be somewhat more than 7 per cent different to cause errors of this kind in a normal teletypewriter with the range finder set in the middle of the range, but if there is some signal distortion other than that from speed discrepancies, such as marking bias, smaller differences in speed would be sufficient to cause errors. Case *j* illustrates the conditions when the sending distributor is slower than the receiving distributor. It will be observed in this case that the first error as the speed discrepancy increased would also be in the fifth impulse as the result of either the fourth impulse being sufficiently prolonged to fall on the fifth selecting position, or the fifth impulse being so late in starting that it is not properly received on the fifth position.

In the illustrations large speed discrepancies have been used so that the shift of the signals could be shown readily on a drawing.

General Transmission Design of *TWX* Network

Telegraph circuits comprising the transmission network employed in teletypewriter exchange service are laid out according to a fundamental plan similar to the toll switching plan [5] used in designing the toll

telephone plant. The teletypewriter switching plan is designed to provide on the most economical basis the circuits necessary for satisfactory connection between any two stations in the country without any special line-up or adjustment of the circuits or apparatus.

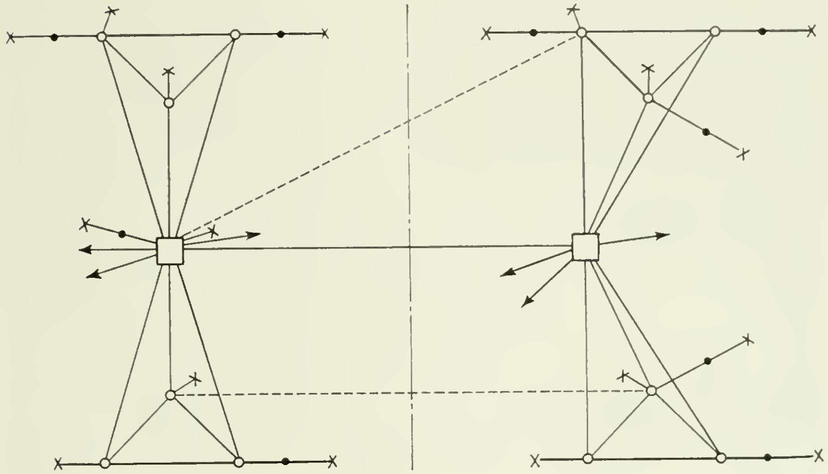
Each switching point has a direct connection to each of the subscriber stations within its area (except for a few stations which are connected to the switchboards by a single channel carrier circuit operating over regular toll telephone circuits when a connection to these stations is desired). In addition it has direct toll circuits to one or more of the other switching points. Eight cities of considerable importance from the standpoint of switching in the national network are designated as "regional centers." These cities, New York, Atlanta, Chicago, St. Louis, Dallas, Denver, San Francisco, and Los Angeles, are interconnected largely by high grade direct circuits and ultimately will be interconnected completely by such facilities. Each of the regional centers has direct circuits to a number of smaller centers designated as "routing outlets" within a given area, which are also interconnected by direct circuits.

The other switching centers, called "teletypewriter centers," which are not required by their position in the networks to handle through business, have direct circuits to one or more routing outlets and may have direct connections to similar nearby centers if traffic justifies it.

The application of the teletypewriter switching plan is illustrated in Fig. 4. Considering only the toll circuits of the basic routes (solid lines connecting switching centers in the figure), it may be noted that within any area where the routing outlets are interconnected by direct circuits, the maximum number of teletypewriter toll lines required for connection between two stations in the area is 3. A very large percentage of the connections can, of course, be made with only one or two toll links. It may also be seen that, assuming all regional centers to be interconnected by direct circuits, a maximum of 5 toll links will serve to connect any two stations in the country, using only the basic routes.

In addition to these basic toll routes, supplementary routes are provided wherever the traffic warrants, as indicated by the dashed lines in the figure. These supplementary routes may be direct circuits between two teletypewriter centers, between a teletypewriter center and a routing outlet or regional center other than that through which it is normally served, or between a routing outlet in one area and a routing outlet or regional center in another regional area. It is obvious from the figure that the effect of these supplementary routes is to reduce the number of toll links and consequently the number of switches involved in certain connections.

The plan permits considerable flexibility with respect to arrangements for future expansion and changes, as growth can be taken care of by the provision of additional switching points or additional direct circuits with practically no change in the fundamental framework.



Regional area *A* Regional area *B*

Fig. 4—Principle of application of teletypewriter switching plan.

□ Regional center
○ Routing outlet
● Teletype center
× Subscriber station
— Basic routes
- - Supplementary routes
→ Routes to other regional centers

Transmission Requirements

In the consideration of the transmission requirements the following items are of importance:

1. The over-all distortion on all connections must be low enough to permit satisfactory service.
2. The distortion on all of the links which will at times be part of built-up connections must be sufficiently low to permit satisfactory transmission when forming a part of such connections.
3. The distribution of distortion between the various toll links and between those links and the subscriber lines should be such as to obtain the desired transmission results with a minimum cost for the plant as a whole.

Transmission Coefficients

The transmission requirements of the over-all connection or of the individual elements are expressed in terms of a system of telegraph transmission coefficients which may be compared roughly to the system of net losses used in telephone transmission work.

In teletypewriter toll circuits of one or more sections the over-all distortion is made up of increments from a number of sources. Experience has shown that in general the over-all distortion of a particular signal element is equal to the algebraic sum of the individual increments. For each specific piece of equipment or element of the circuit the sign and value of the distortion cannot be predicted exactly as they depend upon facts which vary with individual cases. However, representative values of the maximum distortion experienced in a period of moderate length with miscellaneous signals for different types of circuit and equipment may be determined with fair accuracy. Experience and probability theory indicate that the most probable value of the over-all distortion of a telegraph circuit may be computed by taking the square root of the sum of the squares of the corresponding values for the various component parts of the circuit. With this as a basis coefficients have been established for individual telegraph circuits of the various types employed in the *TWX* transmission system. These coefficients are, in general, proportional to the square of the maximum distortion experienced with severe signal combinations under comparatively unfavorable conditions of circuit adjustment, weather conditions, etc., taking into account what is known about the general stability of the particular facility concerned. An estimate may then be made of the transmission impairment to be expected in service with a teletypewriter circuit made up of a number of sections of various types by adding the coefficients of the component parts. For convenience the value of the coefficients has been so chosen that satisfactory operation normally will be obtained over a connection if the sum of the transmission coefficients for the subscriber lines, switchboard circuits, and toll lines involved does not exceed 10.

Using these coefficients the entire transmission system is designed to provide satisfactory transmission between any two subscribers or combinations of subscribers. It is found that subscriber lines less than about 5 miles in length contribute little or no distortion to the over-all connections. Those up to about 35 miles may contribute distortion so as to warrant allowing a coefficient as high as 1.0 or 1.5, and for those up to 50 or 60 miles the coefficient may be as great as 3.5 or 4.0.

The following discussion assumes that the subscriber lines have a coefficient of not more than 1.0 or 1.5 from the subscriber station to the jack connected to the teletypewriter toll line at the switchboard, leaving for the toll links of the connection a maximum coefficient of about 7.0 or 8.0. In the case of intra-area connections involving 3 toll links, a permissible coefficient of 8.0 for all the links of the connections would, of course, permit a coefficient of about 2.7 for each

link. Correspondingly, a connection involving 5 links would permit a coefficient of only 1.6 per link. It happens, however, that the transmission capabilities of the teletypewriter circuits generally in use are such that none of the circuits has a coefficient of less than 2.0 and that single sections of circuit may lie in the range of about 2.0 to 5.0. Practically, the availability of the higher grade circuits is limited by reasons of economy since, for example, the multi-channel carrier telegraph facilities which have coefficients of 2.0 to 2.6 would be too expensive to use for routes where only a few circuits are required or for the shorter links. It is apparent, therefore, that an over-all coefficient of 7.0 to 8.0 cannot be realized on either 4 or 5 link connections without some means for overcoming the over-all distortion. For these cases the operators are provided with connections to regenerative repeaters, which are inserted in series with the circuit and retransmit the teletypewriter signals exactly as they were originally

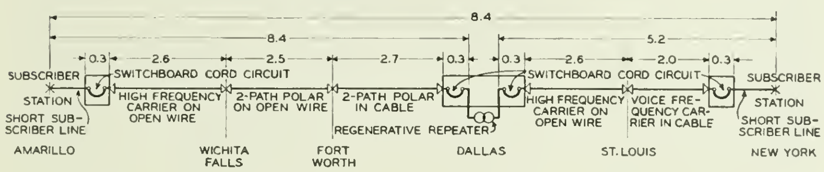


Fig. 5—Diagram of typical teletypewriter exchange service connection requiring regenerative repeater.
Numbers are transmission coefficients.

transmitted into the circuit at the sending end, provided they have not been distorted beyond the point where they can be correctly interpreted by the regenerative repeater. The latter has about the same signal distortion tolerance that a teletypewriter would have if the circuit terminated at that point. Thus the regenerative repeater wipes out the distortion of the preceding toll links and subscriber line so that the coefficient at its output will again be zero. The circuit layout for an actual connection is shown in Fig. 5 illustrating the use of a regenerative repeater.

For the purpose of the teletypewriter exchange circuit layout, it is assumed that regenerative repeaters are available at the switchboards of all regional centers so they may be used to handle 4 and 5-link connections. They are required occasionally at routing outlets to provide satisfactory over-all results on 3-link connections. In the case of 2-link connections the use of regenerative repeaters is ordinarily avoided by limiting the coefficient of the subscriber line, switchboard, and first toll circuit to 5.0. For subscriber lines on which it is not

economical to provide facilities having coefficients as low as 1.5 the traffic routing instructions call for the use of additional regenerative repeaters at suitable points.

The types of telegraph facilities that are used for these various classes of toll link and subscriber line are discussed farther on.

Facilities Used in *TWX* Network—Teletypewriter Stations and Subscriber Lines

A typical teletypewriter station, illustrated in Fig. 6, includes the sending and receiving equipment, together with power supply, and supervisory equipment for initiating a call, informing the attendant of an incoming call, or recalling the switchboard operator during the progress of a connection if desired. These features are described in detail in the paper on switchboards and signaling facilities referred to previously.[1]

Any one of several types of teletypewriter subscriber lines may be used to connect a station with the switchboard from which it is served, the type chosen depending upon conditions in the particular case. A large majority of subscriber lines, however, consist of cable pairs used exclusively for that purpose. In these the telegraph method employed is one in which polar signals (a positive potential for spacing and a negative potential for marking pulses or *vice versa*) are impressed on the subscriber line by a telegraph repeater in the cord circuit at the central office, and neutral signals (the circuit closed for marking and opened for spacing) are transmitted by the sending contacts of the teletypewriter at the subscriber station.

The polar signals transmitted from the central office are symmetrical and the transmission quality of these signals is not affected seriously by the capacitance of the cable loop. As is ordinarily the case in duplex transmission, the current impulses are transmitted differentially through two windings of a relay in the cord circuit repeater which responds to incoming signals but not to the outgoing differentially transmitted signals. To prevent the undesired response of this relay to the outgoing signals, it is necessary that the differential winding not connected to the subscriber line be terminated to ground through an impedance similar to that of the subscriber line. Since subscriber lines from a given type of switchboard are all arranged to use the same current value the resistance component of the station line impedance may be balanced by fixed resistance.

With cable circuits of appreciable length, however, the capacitance becomes of importance. Up to a certain length the effect of the capacitance on balance can be minimized by locating a substantial

portion of the current limiting resistance in series between the subscriber line jack at the switchboard and the subscriber line. For longer circuits an impedance modifying network consisting of capacitance, inductance, and resistance in parallel is inserted in series with the circuit between the subscriber line jack and the subscriber line. The constants of this network are so chosen that the subscriber line

Fig. 6—Typical teletypewriter subscriber station.

will be satisfactorily balanced by the same cord circuit repeater balancing arrangement that is used for the shorter subscriber lines in the office.

At the station the sending contacts and receiving relay or magnet are in series with the subscriber line. Signals from subscriber stations are formed si mply by opening and closing the circuit at the sending

contacts in accordance with the code for the characters being transmitted. When the contacts are closed a current flows in the subscriber line circuit for marking and when they are open this current becomes zero, transmitting a spacing signal.

On long cable pair subscriber line circuits with considerable bridged capacity, the wave shape of the current received in the central office is not symmetrical as regards the marking or spacing conditions, the rate of building up of the marking current being much faster than its rate of decay. This results in marking bias in the received signals. Conversely, in subscriber line circuits containing only series inductance and resistance, the received current builds up gradually to its marking value and decays to zero immediately when the sending contacts are opened for a space. By properly combining the inductance and capacitance, it is possible to produce substantially unbiased signals at the receiving end. In other words, by inserting series inductance in a cable circuit, it is possible to overcome the marking bias effect mentioned above so that practically no distortion occurs in the subscriber line.

The marking bias may also be reduced effectively by the use of series resistance in place of inductance at the subscriber station in cases where it is possible to add a sufficient amount of resistance without reducing the current below the desired value. The effect of series resistance used in this way is to delay the building up of the current when the teletypewriter sending contacts are closed after a spacing signal to compensate for the delay in decay of the received current after the contacts have opened.

Both of the above methods of reducing bias are in use in the present teletypewriter exchange plant. Figure 7 shows the wave shape of uniformly timed marks and spaces received over a 30-mile 19-gauge cable pair, illustrating the effect of the cable capacitance, and the manner in which a wave shaping arrangement, consisting primarily of inductance in this case, reduces the amount of marking bias in the received signal by retarding the building up of current at the start of each marking signal.

Although the majority of subscriber lines are in cables, it is sometimes necessary to serve stations at greater distances from the teletypewriter center or in situations where the use of cable pairs is not practicable. For these, other arrangements must be made. One method of serving such stations is by means of arrangements similar to those of the shorter toll circuits. Generally a telegraph repeater in an office in the vicinity of the subscriber station is used, and transmission between that repeater and the one in the teletypewriter center takes place in the same manner as over a toll circuit of similar length.

Another method for connecting to subscriber stations which cannot be cared for by a metallic cable pair employs a simple telegraph repeater installed as part of the subscriber station equipment. This arrangement as well as the one previously described has the advantage that it permits polar signals to be used in both directions over the subscriber line.

In a few cases which have arisen where telegraph facilities were not readily available between the teletypewriter center and a subscriber station, use has been made of a single-channel voice-frequency carrier

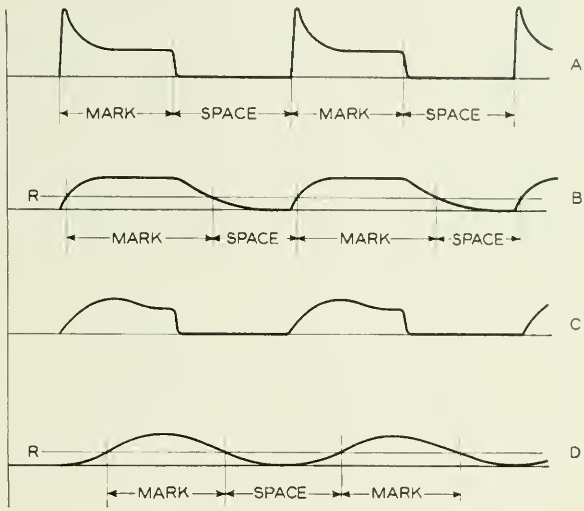


Fig. 7—Effect of wave shaping networks in long subscriber lines operated over cable pairs.

A—Current at subscriber station; no wave shaping network used.
B—Current in A as received at central office.
C—Current at subscriber station; loop equipped with wave shaping network.
D—Current received at central office; loop equipped with wave shaping network.
R—Current required to operate receiving relay of repeater in central office.

telegraph arrangement by means of which the transmission takes place over standard telephone circuits. A small carrier telegraph terminal arrangement is mounted on the back of the teletypewriter table, and a corresponding carrier terminal is located in the teletypewriter center in a trunk circuit between the teletypewriter switchboard and the telephone toll board. Special operating procedures are set up so that whenever the subscriber initiates a call, connection is established by telephone operators over telephone circuits to the above mentioned carrier trunk circuit at the teletypewriter center, and the teletypewriter switchboard operator is notified

of the call and given the number of the subscriber by whom it is made. From the subscriber's standpoint calls are made with this equipment in practically the same manner as when ordinary telegraph facilities are employed.

Switchboards

The switchboards used in teletypewriter exchange service contain facilities for interconnecting subscriber lines, connecting them with toll circuits, or interconnecting toll circuits as required, together with the necessary means for establishing and supervising the connections. They are described in considerable detail in the previously referred to paper on switchboards and signaling facilities.[1] As indicated in the discussion of subscriber lines, the transmission circuit through the switchboard is essentially a differential duplex telegraph repeater. One such repeater is connected between the cords of each pair. This repeater is so designed that it introduces very little distortion in the connection. The coefficient of the switchboard cord circuit is 0.3. Figure 8 is a schematic diagram showing the principle of the transmission circuit.

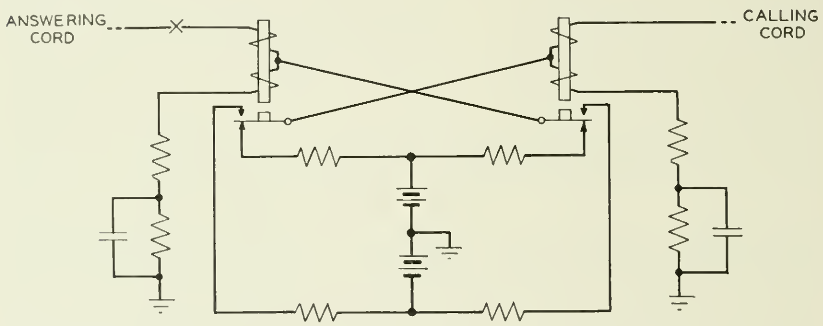


Fig. 8—Principle of switchboard transmission circuit.
Operator's teletypewriter inserted at X when required.

Toll Circuits

The toll circuits of the teletypewriter exchange network are of the standard types that are in general use for telegraph transmission. These include voice-frequency carrier telegraph systems on cable circuits [6] or on channels of carrier telephone circuits on open-wire lines, high-frequency carrier telegraph systems on open wires,[7] metallic systems on cables,[8] and two-path polar and differential-duplex grounded telegraph circuits.[9] An idea of the relative capabilities of these types of facilities may be obtained from Table I which shows the coefficients of a single section of each type.

TABLE I

TRANSMISSION COEFFICIENTS FOR 60-SPEED TELETYPEWRITER EXCHANGE CIRCUITS

Type of Circuit	Coefficient per section *	Maximum Section Length Normally Used, Miles
D-C. grounded system on open wire	2.5 to 4	300
D-C. metallic system on cable circuits	2 to 3	150
High frequency carrier system on open wire	2.6	1,150
Voice frequency carrier system on cable or open wire circuits	2.0 to 2.2	3,500

From the coefficients given in the table and the earlier discussion of the teletypewriter switching plan, it is apparent that the carrier systems, voice-frequency or high-frequency, where available, are most suitable for the longer backbone toll circuits of the nationwide network. For the short circuits of from 100 to 200 miles where cable plant is available, the metallic telegraph circuits on cable are extensively used, while for the scattering circuits of similar length, and most of the shorter toll circuits, use is made of two-path polar and differential duplex facilities. In some instances where single section facilities of the required grade are not available between two centers, regenerative repeaters permanently associated with multi-section circuits are used to provide satisfactory over-all circuits. Also in certain instances where circuits are not required for through switching, multi-section circuits without regenerative repeaters are sometimes provided and classified "for terminal purposes only."

All the components of the network—teletypewriters and their associated subscriber lines, transmission circuits in the switchboards, and toll circuits interconnecting the switchboards—are designed to give a satisfactory over-all transmission performance with a minimum cost for the plant as a whole. Results obtained in service indicate that the system is meeting a commercial need and that its performance is satisfactory, but developments are continually under way to effect further improvements in service and economies in operation as experience is gained with the system.

REFERENCES

1. "Switchboard and Signaling Facilities of the Teletypewriter Exchange System," A. D. Knowlton, G. A. Locke, and F. J. Singer, this issue of the *Bell System Technical Journal*, and *Electrical Engineering*, v. 55, September 1936, pp. 1015–25.

* The term "section" as here used designates the part of a telegraph circuit between 2 telegraph repeaters or a section of a telegraph circuit without any intermediate telegraph repeaters. For example, a telegraph repeater section operated by the voice-frequency carrier telegraph method is that part of the circuit between carrier telegraph terminal sets, regardless of the number of intermediate telephone repeaters in the carrier circuit.

2. "Certain Topics in Telegraph Transmission Theory," H. Nyquist, *A. I. E. E. Trans.*, v. 47, 1928, pp. 617–44.
3. "Certain Factors Affecting Telegraph Speed," H. Nyquist, *A. I. E. E. Jour.*, v. 43, February 1924, pp. 124–30, and *Bell Sys. Tech. Jour.*, v. 3, April 1924, pp. 324–46.
4. "Measurement of Telegraph Transmission," H. Nyquist, R. B. Shanck, and S. I. Cory, *A. I. E. E. Jour.*, v. 46, March 1927, pp. 231–40.
5. "General Switching Plan for Telephone Toll Service," H. S. Osborne, *A. I. E. E. Trans.*, v. 49, 1930, pp. 1549–57, and *Bell Sys. Tech. Jour.*, v. 9, July 1930, pp. 429–447.
6. "Voice Frequency Carrier Telegraph System for Cables," B. P. Hamilton, H. Nyquist, M. B. Long, and W. A. Phelps, *A. I. E. E. Jour.*, v. 44, March 1925, pp. 213–18.
7. "Carrier Current Telephony and Telegraphy," E. H. Colpitts and O. B. Blackwell, *A. I. E. E. Jour.*, v. 40, 1921, April, pp. 301–15, May, pp. 410–21, and June, pp. 517–26.
8. "Metallic Polar-Duplex Telegraph System for Cables," John H. Bell, R. B. Shanck, and D. E. Branson. *A. I. E. E. Trans.*, v. 44, 1925, pp. 316–25.
9. "Modern Practices in Private Wire Telegraph Service," R. E. Pierce, *A. I. E. E. Trans.*, v. 50, 1931, pp. 426–35.

A New Telephotograph System *

By F. W. REYNOLDS

Transmission of photographs over telephone wires was begun commercially several years ago, but recent improvements have increased to 11 by 17 inches the size of photograph that could be transmitted and have made it possible for the picture to give much more information. The new machines used for sending and receiving photographs are described in this paper, and the requirements and control of the wire system necessary to prevent imperfections in the picture and to permit switching of sending and receiving stations are discussed.

A TELEPHOTOGRAPH message service between New York, Chicago, and San Francisco was initiated in April 1925 by the Bell System, and was extended during the following two years to five additional cities. Experience in the operation of this service, using equipment previously described,[1] indicated that a number of improvements were desirable in order to meet more satisfactorily the apparent requirements of this form of communication. Development work was undertaken to effect these improvements, and this paper describes the new equipment and some of the features involved in establishing a leased wire telephotograph network connecting 26 cities as shown in Fig. 1.

During the eight years of operation of the first Bell System telephotograph service the performance of the system was observed, analyses made of the material transmitted, and opinions formulated regarding the acceptability of the received pictures. The early equipment required the preparation of the material for transmission as a film transparency in an area not exceeding $4\frac{1}{4}$ inches by $6\frac{1}{2}$ inches. This relatively small image field combined with the use of 100 scanning lines per inch and the added photographic operations to prepare the material for transmission were considered as limiting the usefulness of this new service. For example, in transmitting many of the forms of printed matter it was necessary to divide the copy into overlapping sections, to transmit each piece separately and to assemble the sections as a composite picture at the receiving point. Obviously this procedure could not be applied advantageously to a photograph or news picture and therefore the maximum information content of such transmissions was limited by the small size of image field and the

* Published in *Electrical Engineering*, September 1936. Presented at A. I. E. E. Southwest District meeting, Dallas, Texas, October 26–28, 1936.

[1] For all numbered references see list at end of paper.

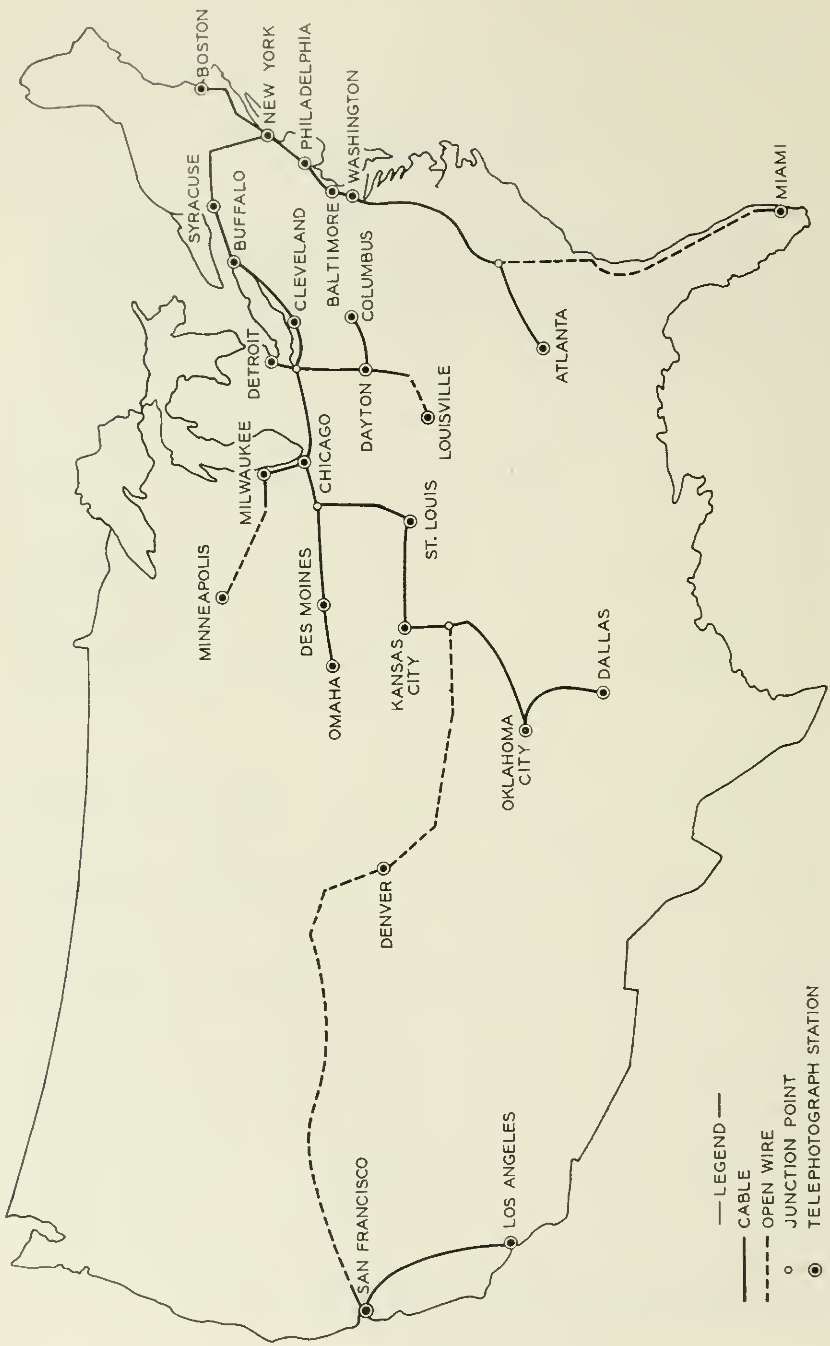


Fig. 1—A leased wire telephotograph network in the United States.

number of scanning lines employed. Certain types of pictures such as portraits, small groups, and others of a rather limited information content were transmitted satisfactorily with this early equipment, but transmissions of those pictures containing much greater amounts of information frequently were regarded as inadequate.

In formulating specific requirements for the new telephotograph system consideration also was given to the increasing interest in news pictures and to the trend in this country toward improvement of newspaper halftone reproduction. The former public demand for pictures of the occasional catastrophe or outstanding news event is today apparently being supplemented by an interest in the pictorial reporting of even minor news items. These factors are reacting to elevate the standards for acceptable telephotographs to a plane where newspaper halftone reproduction of original and transmitted pictures may soon be comparable in quality and information content. The requirements met by the new telephotograph system are summarized briefly in the following paragraphs.

Scanning

Pictures are scanned by reflected light at 100 lines per inch. This permits direct transmission from original subject matter in the majority of cases without recourse to special preparation such as photographic copying.

Size of Image Field

A useful image field is provided for scanning and reproducing pictures of various sizes up to and including 11 inches by 17 inches. This area is sufficient to accommodate most sizes of subject matter likely to be encountered in telephotography and is well adapted to transmission of black and white information such as financial statements, advertising layouts, and the like. Furthermore, it provides a practical method of varying the information content of received pictures by using original prints of appropriate sizes. This point is illustrated in Figs. 9 and 10, which are reproductions from transmissions made from prints of the same subject which were respectively 4¼ by 6⅛ inches and 10 by 14½ inches. The useful circumference of the picture cylinder employed is 11 inches. In the case of news pictures, which are ordinarily distributed as 8 by 10 inch photographic prints, the remaining one-inch space on the circumference of the cylinder may be utilized for transmitting the caption as part of the picture.

Speed of Transmission

The image field in the new equipment is scanned at 100 lines per inch with a velocity of 20 inches per second, which results in the

transmission of one inch of picture per minute, measured along the axis of the picture cylinder. This rate of scanning produces essential signal frequencies extending approximately from zero to 1,000 cycles per second and is more than double the speed of transmission used in the earlier equipment. However, by employing the single-side-band method of transmission it has been possible to use this speed of transmission over telephone circuit facilities of normal band width but specially modified as described in a later section.

Synchronism

Operation of the earlier Bell System telephotograph equipment over long telephone circuits indicated the desirability of providing improved means for synchronizing the sending and receiving equipment. Accordingly, development work was undertaken, and local frequency sources of the required stability were made available to permit independent speed control without transmitting synchronizing signals. Experimental oscillator units were installed for tests at three telephotograph stations about two years after the opening of the public telephotograph service in 1925. Experience gained from the use of these oscillators, which were vacuum tube driven tuning forks maintained within close temperature limits, indicated that this method was practicable, although the particular arrangements employed at that time could be advantageously improved.

A new design of tuning fork controlled oscillator has been provided in the new equipment whose frequency can readily be adjusted and maintained constant to within a few parts in a million. This difference in speed between sending and receiving machines is so slight that skewing of the received picture is not noticeable.

Starting and Phasing

The simultaneous starting of all machines participating in the transmission and reception of a picture is effected by means of a signal sent over the line by the transmitting machine. Phasing of the machines is automatic, since all are started simultaneously from the same angular position by a positive action clutch. This requirement is similar to that met by the earlier equipment, but more difficult to fulfill because of the use of a much larger picture cylinder. It required the development of a new type of clutch which would permit a gradual increase in velocity of the cylinder and yet be positive in action. The fulfillment of this requirement is important as it assures accurate phasing without consuming valuable circuit time, irrespective of the number of machines involved in a transmission.

Design

In addition to meeting the above general requirements the new design includes arrangements for daylight operation, a new type of driving motor, and scanning with a pulsating beam of light whereby the photoelectric current can be amplified by a-c. methods.

Description of the New Telephotograph Equipment

The general specifications outlined in the preceding paragraphs are embodied in the new telephotograph equipment now being manufactured. Telephotograph equipment of this type for a station arranged to send and receive pictures consists of a sending machine and a receiving machine mounted on separate tables (see Figs. 2 and 3),

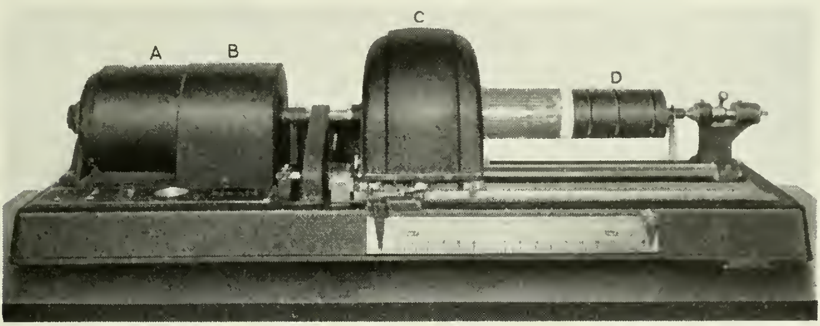


Fig. 2—Telephotograph sending machine.

A—Motor.
B—Clutch.
C—Optical system.
D—Picture cylinder.

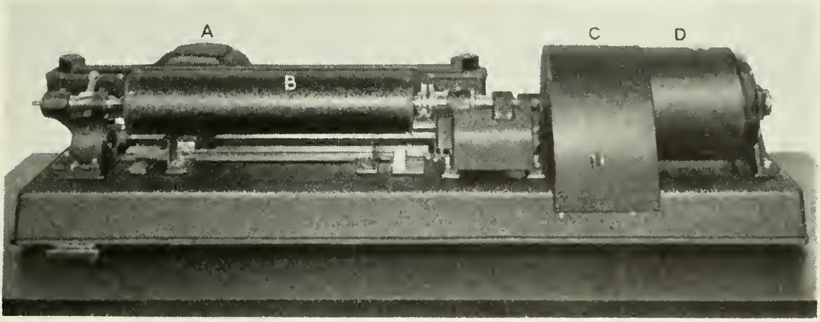


Fig. 3—Telephotograph receiving machine.

A—Optical system.
B—Cylinder housing.
C—Clutch.
D—Motor.

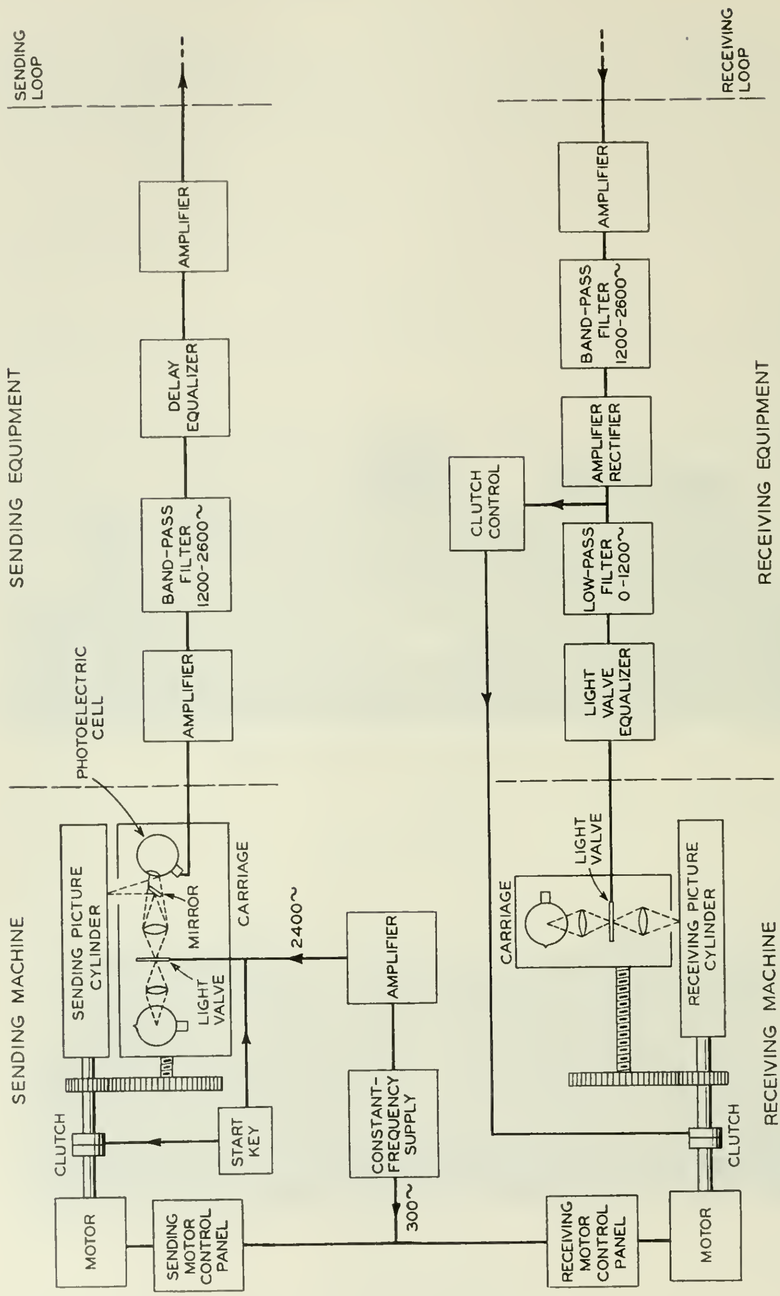


Fig. 4—Schematic diagram of sending and receiving equipment for one station.

two bays of relay-rack-mounted apparatus, and a cabinet of power supply equipment. A third bay comprising loop terminating arrangements, telephone, and loud speaker equipment may be furnished by the telephone company if ordered by the customer. This telephotograph equipment connected by suitable circuits will transmit pictures and other forms of graphic information from point to point or from one to a number of points simultaneously.

Figure 4 is a schematic diagram illustrating the functional relationships of the various units of this equipment. Certain features of these units that may be of special interest have been selected for description in the following.

Motor and Associated Speed Control Circuit

Although the driving motor for the telephotograph machine is essentially of the d-c. shunt type, it functions in combination with its associated speed control equipment as a synchronous unit and upon starting locks automatically in synchronism with the frequency generated by the local carrier and motor control oscillator. This is accomplished in a manner similar to that previously used in television equipment demonstrated by the Bell System.[2, 3] An inductor type generator built into the frame of the motor delivers an a-c. output of 300 cycles per second at the normal speed of the motor, 100 r.p.m. The output of the generator is impressed upon the plates of two vacuum tubes the grids of which are energized by the 300-cycle output of the carrier and motor control oscillator as shown in Fig. 5.

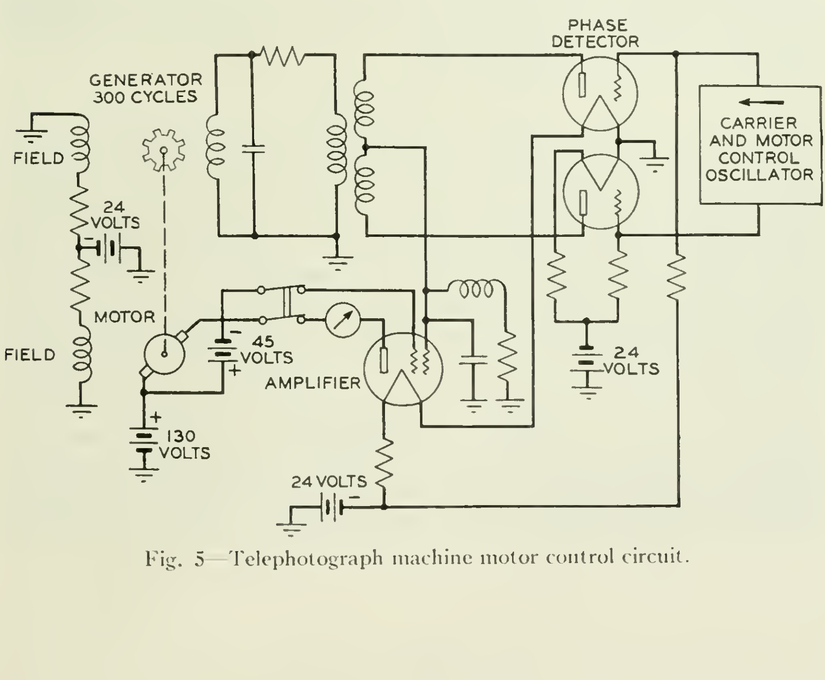


Fig. 5—Telephotograph machine motor control circuit.

These tubes act as a phase detector and vary the input voltage across an amplifier which supplies the total armature current of the motor. Armature rather than field control is employed to obtain faster and more complete regulation. The capacitor across the terminals of the generator armature, tuning the circuit to a frequency slightly in excess of 300 cycles per second, and the coupling impedance between phase detector and amplifier acting as a low-pass filter, assist in preventing hunting of the motor.

Clutch

Connection between driving motor and picture machine is made through a positive action clutch electrically operated. This clutch gradually applies the driving torque to the picture machine during the starting interval. It operates on the principle of storing energy in a coiled spring during the first part of the starting interval while the velocity of the machine is increasing and then allowing this energy to be released gradually by an escapement mechanism while the parts of the clutch are assuming their normal operating position. The time interval required for complete operation of the clutch corresponds to three or four revolutions of the picture cylinder but variations in the length of this interval do not affect the accuracy of phasing, inasmuch as the latter is determined by the time of operation of the clutch trip magnet and each receiving machine may be readily adjusted to compensate for this variation.

Circuit arrangements associated with the clutch of the receiving machine permit its operation from a starting signal received over the line from the transmitting machine.

Sending Optical System

The optical system of the sending machine is arranged to direct a scanning light beam upon the surface of the picture which is mounted on a cylinder. This scanning beam, attenuated by reflection from the various shades of the picture, is directed to a photoelectric cell. The illumination is obtained from a small incandescent lamp and is interrupted in passing through the aperture of a double ribbon light valve. This double ribbon light valve, which is a modification of a type previously described,[4] is actuated by the picture carrier frequency, 2,400 cycles per second, and its interruption of the scanning light beam permits the use of a-c. methods of amplification of the photoelectric currents. Aside from its general simplicity and freedom from the usual difficulties experienced with rotating light choppers, this type of interrupter readily effects a sinusoidal variation in illumination.

It is obvious that, since the illumination incident upon the picture is pulsating at the carrier frequency, the currents present in the output of the photoelectric cell will consist of the picture signal currents and the carrier frequency modulated by these currents, the picture itself acting as a simple direct product modulator.

Filters and Delay Equalizer

The application of single-side-band transmission methods to the present telephotograph equipment has resulted in the design of electrical filters of rather unusual phase shift and attenuation-frequency characteristics. It has previously been pointed out in connection with a discussion of telegraph transmission theory [5] that three conditions should be fulfilled for single-side-band transmission:

1. The system should have a linear phase shift-frequency characteristic.
2. The sluggish in-phase component of the signal resulting from a displacement of the carrier from the middle of the transmitted band should be eliminated.
3. The received quadrature component resulting from the loss of the component of the side band suppressed, equal in magnitude but opposite in sign, should also be eliminated.

The first two conditions are met by the careful design of a delay equalizer network and a special filter giving a suitably shaped admittance characteristic for the system. This characteristic exhibits a type of symmetry about the carrier frequency which would result in a superposition of the regions adjacent to the carrier if rotated about this point. Consequently the attenuation of the filters and associated delay equalizer should be 6 decibels greater at the carrier frequency than at the middle of the band of the transmitted frequencies, in addition to meeting the requirement of a linear phase shift-frequency characteristic. Over-all attenuation and phase shift-frequency characteristics of the filters and equalizer of the sending and receiving equipment of the present design are shown in Fig. 6.

In regard to the third condition for single-side-band transmission, experiments have shown that the effect in received pictures of the quadrature component is not of practical importance in the present equipment. The quadrature component is determined essentially by the slope of the signal envelope which is in turn restricted by the equivalent transfer admittance characteristic [6] of the scanning aperture, and by the slope of the filter characteristic to meet condition 2.

Receiving Optical System

The receiving optical system of the new telephotograph equipment is similar in its general aspects to that employed in the earlier Bell System equipment. Illumination from an incandescent lamp is

directed to the receiving photographic emulsion through the aperture of a single ribbon light valve. The latter, however, is operated by the rectified picture currents instead of by the modulated picture carrier current as used in the earlier equipment. This change results in very simple yet efficient optical arrangements for receiving a

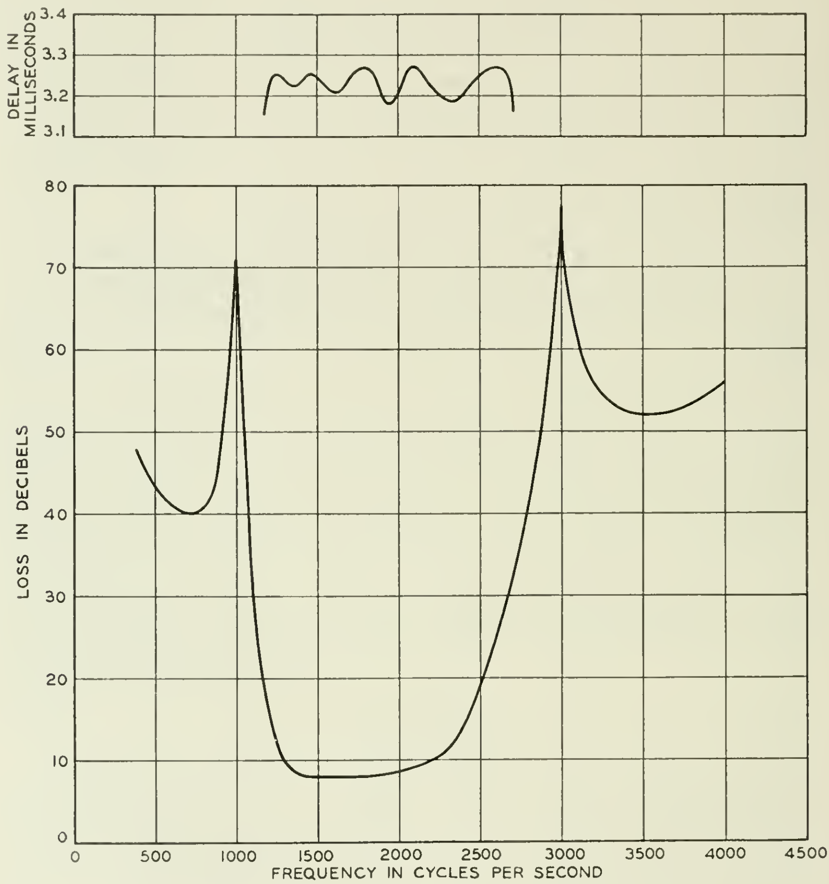


Fig. 6—Over-all characteristics of filters and delay equalizer.

variable density constant line width picture with no apparent structure. The aperture of the light valve, which is uniformly illuminated by the incandescent lamp, is adjusted so that the width of its image on the receiving emulsion is 0.01 inch. The height of the aperture, which determines the exposure, is regulated by the instantaneous position of the light valve ribbon and this is proportional to the received picture

currents. A uniformly illuminated field is obtained at the light valve aperture with a minimum loss of light by using a sphero-cylindrical condenser lens which focuses the diameter of the helical filament of the lamp without imaging individual turns of the helix at the plane of the aperture. Imaging of the lamp filament with the usual type of spherical condenser lens would result in non-uniformity of illumination not only because interstices between individual turns of the helix have an intrinsic brilliancy much greater than the outer surface of the filament but also because of the angular variation of the masking effect of the turns of the helix upon the illumination emerging from the interstices. The ribbon of the light valve is tuned mechanically to resonance at 1,200 cycles per second, and is shunted by an equalizer [7] consisting of inductance, capacitance, and resistance in series which is tuned to the resonant frequency of the ribbon, thereby producing a flat response-frequency characteristic over the useful range of signal frequencies.

Carrier and Motor Control Oscillator

This portion of the equipment furnishes the carrier frequency of 2,400 cycles per second and the motor control frequency of 300 cycles per second accurate to within a few parts in a million. The arrangements used consist of a 300-cycle tuning fork within a temperature regulated container, a vacuum tube amplifier circuit designed to provide controlled regenerative operation of the fork, and a vacuum tube harmonic generator for supplying the carrier frequency.

Although this general method for obtaining a constant frequency is old and has been described previously,[8, 9, 10] in view of its importance in the operation of the present telephotograph equipment it may be of interest to indicate briefly the specific arrangements employed.

The tuning fork is made of a heat treated nickel chromium steel alloy to obtain a small frequency-temperature coefficient and is mounted in a thermostatically controlled metal cylinder wound with a heating coil over which are wrapped alternate layers of copper and felt to provide attenuation of heat transfer.[11] The pick-up and drive coils associated with the fork are connected to the vacuum tube amplifier circuit as shown in Fig. 7. The frequency of a fork is affected by a number of factors including temperature, amplitude of vibration, and aging of the material. Since it is impracticable to maintain constant all of the factors involved, it is necessary to provide means for occasional adjustment to meet the requirements for constancy desired in picture transmission. In the present equipment the temperature of the fork is maintained within ± 0.1 degree of its nominal value of 50 degrees centigrade; two adjustments are provided

for changing the amplitude, one of which varies the grid potential of a vacuum tube which acts to limit the current supplied the driving coil, and the other, a variable capacitor in the circuit containing the

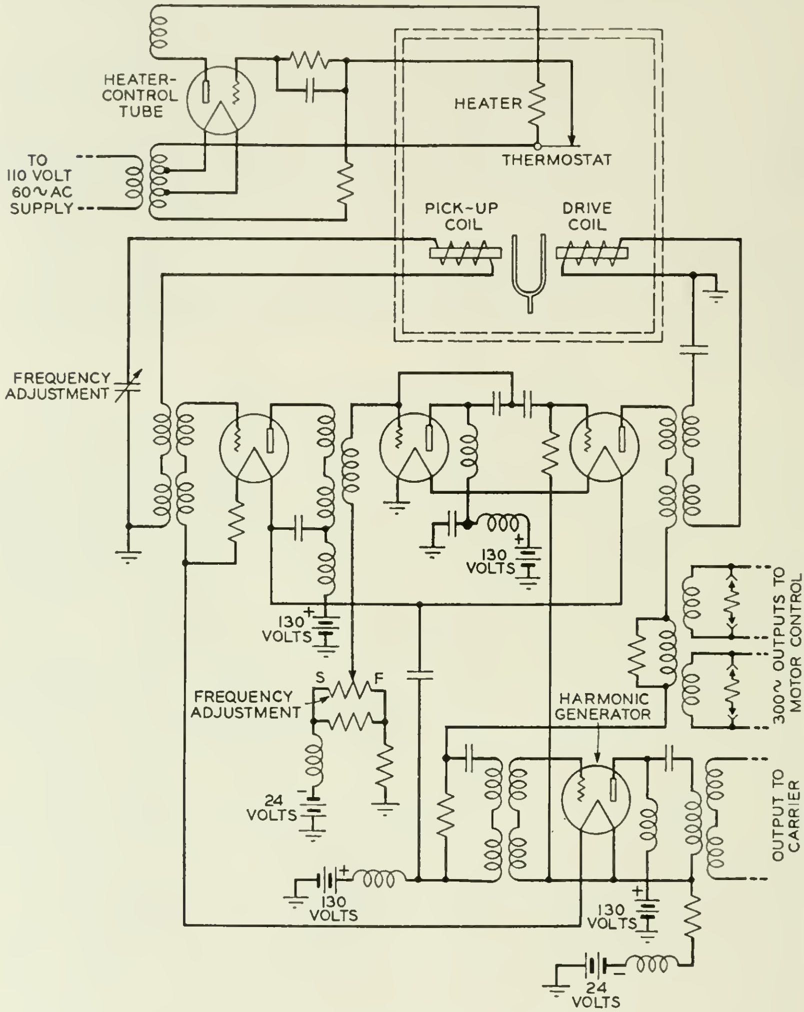


Fig. 7—Carrier and motor control oscillator.

pick-up coil varies the phase relation between the currents in the drive and pick-up coils. Three outputs are provided from the oscillator, two for the sending and receiving motor control circuits and the third for the carrier frequency supplied the sending light valve.

All of these outputs terminate in high impedance circuits and have no appreciable reaction upon the constancy of operation of the fork.

Line Facilities Used with the New Telephotograph Equipment

Requirements for the communication channel used in the transmission of pictures are obviously dependent upon the characteristics of the telephotograph equipment employed and the amount of degradation resulting from transmission which can be tolerated. In general, telephotograph equipment capable of recording the transmitted signals with a degree of fidelity of the order required for good pictures may also record certain extraneous disturbances in the transmission channel which will appear as blemishes on the received picture. The more important of these disturbances are abrupt variations in line net loss, delay distortion, certain types of noise, echoes, and crosstalk. With the exception of delay distortion which is more pronounced for the new equipment because of its higher speed of transmission, the requirements relating to the other disturbances are comparable to those applying to the earlier Bell System equipment. Experience over a period of years with the earlier equipment indicated that selected telephone circuits, specially conditioned to adapt them to picture transmission and established as a regular network, could be relied upon to give consistently good results. This general procedure has been followed in establishing wire networks for use with the new telephotograph equipment and unretouched reproductions of typical news pictures received over such circuits are shown in Figs. 8 and 11.

The circuit facilities employed with the new telephotograph system are 4-wire H-44-25 side circuits in cable [12] where available and elsewhere 2-wire open-wire [13] side and physical circuits.* These facilities are provided with delay equalizer networks for the frequency range from 1,200 to 2,600 cycles per second and precautions are taken to minimize various transmission disturbances. Means are provided, controlled by the sending telephotograph equipment, to prevent operation of the transmission regulating network relays on cable circuits during the transmission of a picture, and to obtain one-way transmission over 2-wire circuits. A wire network of nearly 8,000 miles established as outlined above and connecting 26 stations of the new telephotograph equipment has been in operation for more than a year, giving reliable and technically satisfactory service.

* A side circuit is a physical circuit that is used for one of the paths of a phantom circuit; the notation H-44-25 indicates a loading coil spacing of 6,000 feet, inductance of physical or side circuit loading coils 44 millihenries, and inductance of phantom circuit loading coils 25 millihenries.

Transmission Requirements

The effects of extraneous line disturbances may or may not be particularly objectionable in a specific case, depending upon their magnitude and form and also on the nature and use of the received picture. Furthermore, the predominance of the recorded disturbance may also be affected by normal variations in the adjustments of the telephotograph equipment. It is not practicable, therefore, to establish precise limits for the transmission requirements. The following values are mentioned as illustrative of the order of magnitude for some of the more important requirements applying to circuits used with the new telephotograph equipment, and which experience has shown will give generally satisfactory results.

(*a*) *Line Net Loss*

Abrupt variations in line net loss of 0.2 decibel or greater usually will produce a noticeable change in shade of the received picture. However, a gradual variation in net loss occurring over a period of minutes is less objectionable and in many instances a change of as much as 2 or 3 decibels during a transmission can be tolerated.

(*b*) *Noise*

Noise of a single-frequency type is likely to be recorded in the received picture as an objectionable *moiré* pattern if the difference between the maximum signal and interference energy is less than 50 decibels. However, if the interference energy is distributed over a relatively wide frequency band an energy difference of about 35 decibels usually can be tolerated.

(*c*) *Delay Distortion*

Delay distortion introduced by the circuit, if of sufficient magnitude, may produce multiple outlines along the edges of objects or lines in the received picture and result in a loss or general masking of picture detail. In order that this effect may be inappreciable in pictures received with the new telephotograph equipment it is desirable that the maximum deviation in envelope delay throughout the useful frequency band (1,200 to 2,600 cycles per second) be less than $\pm$ 300 microseconds.

D-C. Control Circuit

Sudden small variations in line net loss are normal on toll cable circuits in the United States as the result of the stepping of the regulating network relays, which, under control of a pilot wire regulator, compensate for the effect of temperature changes on the attenuation

Fig. 8—Telephotograph received at New York from Miami. Size of received picture was 7¾ by 13½ inches. Reproduced by courtesy of the Associated Press.

Fig. 9—Telephotograph. Received 4¼ by 6⅛ inches with 100 lines per inch. Reproduced by courtesy of the Associated Press.

Fig. 10—Telephotograph. Received 10 by $14\frac{1}{2}$ inches with 100 lines per inch. Reproduced by courtesy of the Associated Press.

Fig. 11—Telephotograph received at New York from Baltimore. Size of received picture was $8\frac{3}{4}$ by $14\frac{1}{2}$ inches. Reproduced by courtesy of the Associated Press.

of the circuit. Since these sudden variations in net loss produce noticeable changes in shade of the received picture, means similar to those employed with the earlier Bell System telephotograph equipment have been made available to prevent these relays from operating while a picture is being transmitted. Simple types of control units actuated by signals transmitted over a control circuit are connected to each regulating repeater associated with the picture circuit. This control circuit consists of two one-way d-c. channels obtained by compositing the telephotograph circuit and extended to each telephotograph station over simplexed loop arrangements. The control circuit is also arranged to perform other functions such as effecting one-way transmission of the 2-wire circuits during a picture transmission. The operation of the control circuit normally is performed automatically at the sending telephotograph station.

Inasmuch as the transmission requirements for this control circuit are very lenient compared with those for telegraphy, it has been possible to employ simple types of d-c. repeaters as illustrated in Fig. 12. A signal from the subscriber's sending equipment operates the receiving relay of the station repeater, which in turn places a ground on the M lead and thus transmits the signal to all line repeaters which may be associated with this junction. Only one direction of operation at a time is possible so that when a sending telephotograph station takes control at the beginning of a picture transmission the control circuit is operated and remains in this condition until released automatically at the end of the transmission. A slow release circuit is provided in the d-c. repeater used at regulating network points on the cable circuits and also in another type of repeater, not shown but used on open-wire circuits to obviate false operation of the repeaters as the result of interruptions of less than two seconds duration.

Delay Equalization

Delay equalization [14] of telephotograph circuits is not new, but was applied in 1925–26 to certain medium-heavy loaded toll cable circuits between New York and Boston which were used in the early Bell System telephotograph service. (This application was discussed in reference 14 relative to delay distortion, and examples of transmitted printed matter were reproduced.) However, because of the increased speed of transmission of the new telephotograph equipment and the demand for longer circuits for picture transmission it has been necessary to make further application of delay equalization to some of the more common types of circuits used for this purpose.

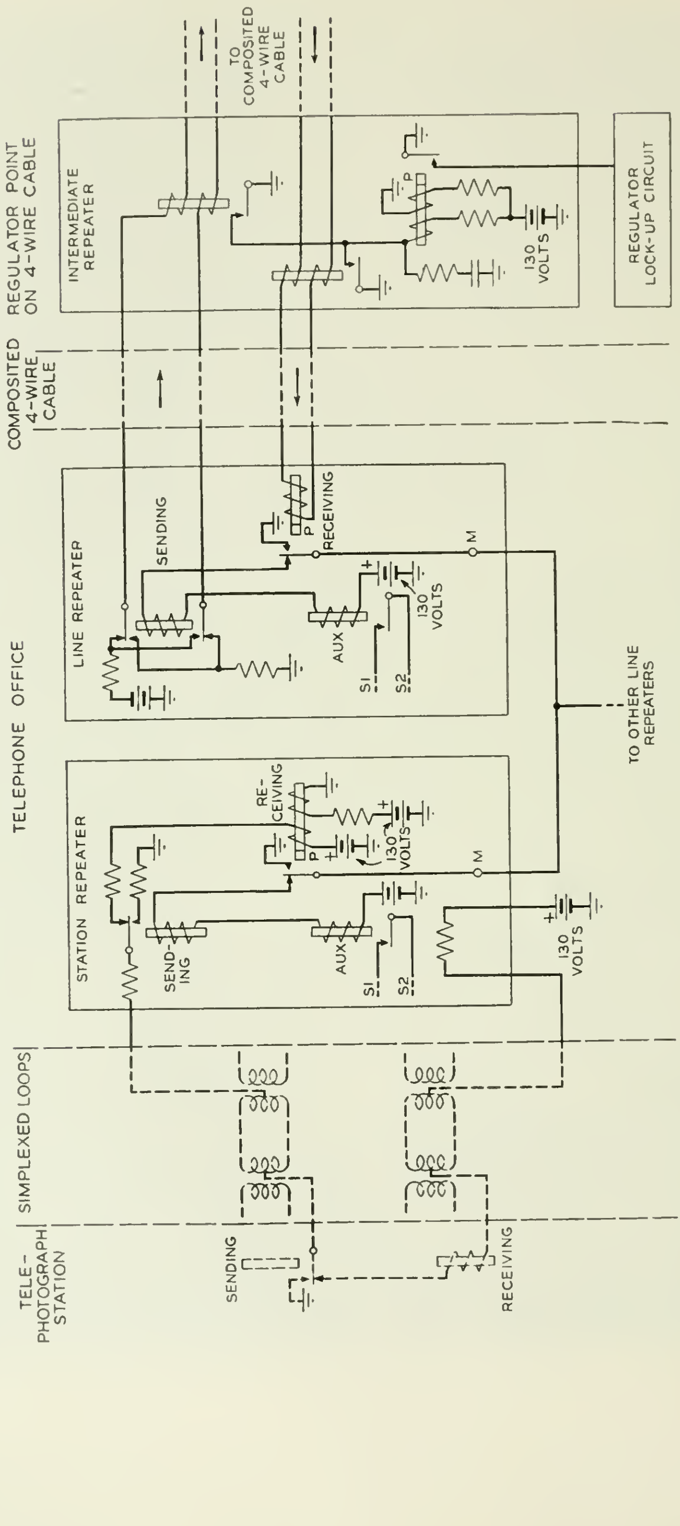


Fig. 12—D-c. control circuit repeaters.

Delay distortion in H-44-25 cable circuits, which is largely the result of the loading, has been compensated by delay networks consisting of a basic unit correcting for 150 miles of composited 19-gauge side circuit, adjustable in 10-mile steps, and a "mop-up" unit of four sections for more complete compensation. A balanced lattice type of structure was used in the design of these equalizers. Figure 13

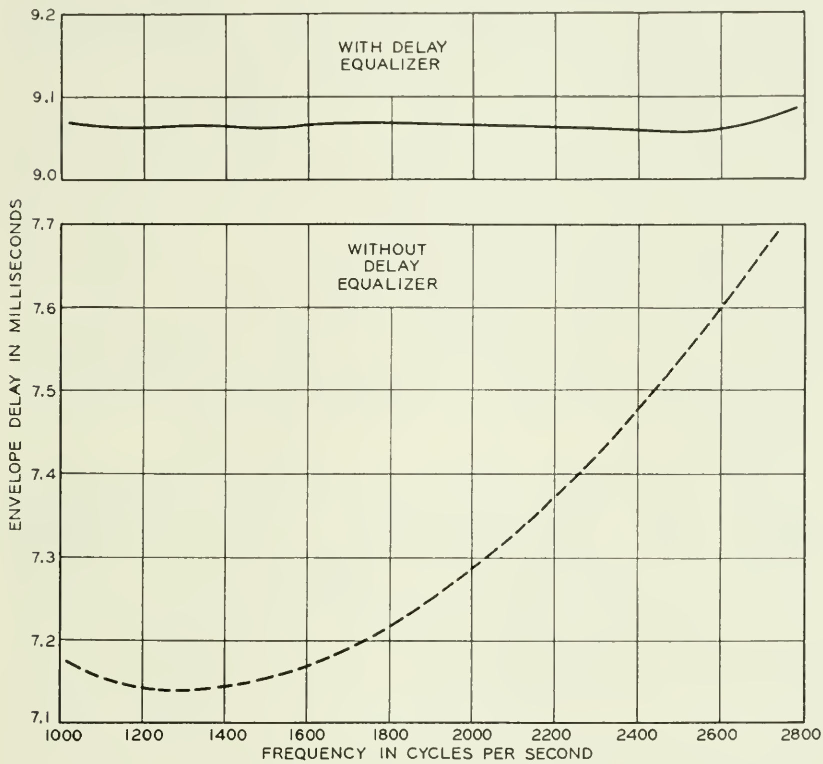


Fig. 13—Delay characteristic of 135 miles of H–44 repeatered and composited side circuit before and after equalization.

illustrates the application of equalizer units to cable circuits and shows the delay characteristics before and after equalization. Similar types of delay equalizers have been applied to open-wire circuits, in which case it is the equipment located at the repeater stations rather than the line itself which is responsible for delay distortion. The delay equalizers are normally located at terminal and bridging points on the telephotograph network and at some intermediate points such as junctions of open-wire and cable circuits.

Telephotograph Networks

One of the obvious advantages of network operation of telephotograph stations is that it offers a means for rapid and simultaneous distribution of facsimile information and pictures to a large number of receiving points. This method of operation appears to be particularly advantageous for use by the large news-picture gathering and distributing agencies giving a nation-wide service. Such network operation of a number of telephotograph stations presents additional requirements, not mentioned in the preceding paragraphs, which may be of general interest.

Requirements encountered in connecting a large number of sending and receiving stations were that any sending station should be able to transmit a picture simultaneously to all receiving stations, and that any one station could be selected as the transmitting point, establishing a new direction of transmission with a minimum loss of time. The situation has been met by permanently bridging each telephotograph station, consisting of separate sending and receiving equipment, to the wire network on a 4-wire basis using separate sending and receiving station loops and performing automatically such switching operations as may be involved in altering the direction of transmission.

Typical arrangements which have been used at a bridging point on a telephotograph network are illustrated in Fig. 14. Suppose, for example, that the telephotograph station at this point wishes to transmit a picture to the network. Operation of a key associated with the subscriber's telephotograph transmitting equipment sends out a d-c. signal over the simplexed loop to the control circuit station repeater at the local telephone office. Since this repeater is multipled with the d-c. repeaters associated with each of the telephotograph circuits connected at this point, the signal is transmitted over the entire network and the switching operations performed to place the circuits in condition to send a picture from this point. The d-c. repeaters at the local telephone office also cause short circuits to be applied to the incoming transmission paths which are connected to the bridging networks, thus preventing the temporarily inactive parts of the circuit from contributing possible disturbances to the outgoing paths being used. This figure also indicates the switching operations performed on the 4-wire terminating set. At the conclusion of the transmission the d-c. control circuit is automatically released by the transmitting machine and the circuits returned to the initial two-way condition permitting any station on the network to seize control of the circuits for picture transmission. Signal lamps are provided at

all d-c. repeater points and are actuated by the d-c. control circuit to indicate when pictures are being transmitted over the network and also the direction of transmission.

The problems involved at junction points in connecting a number of circuits, particularly of the 4-wire type, have been simplified

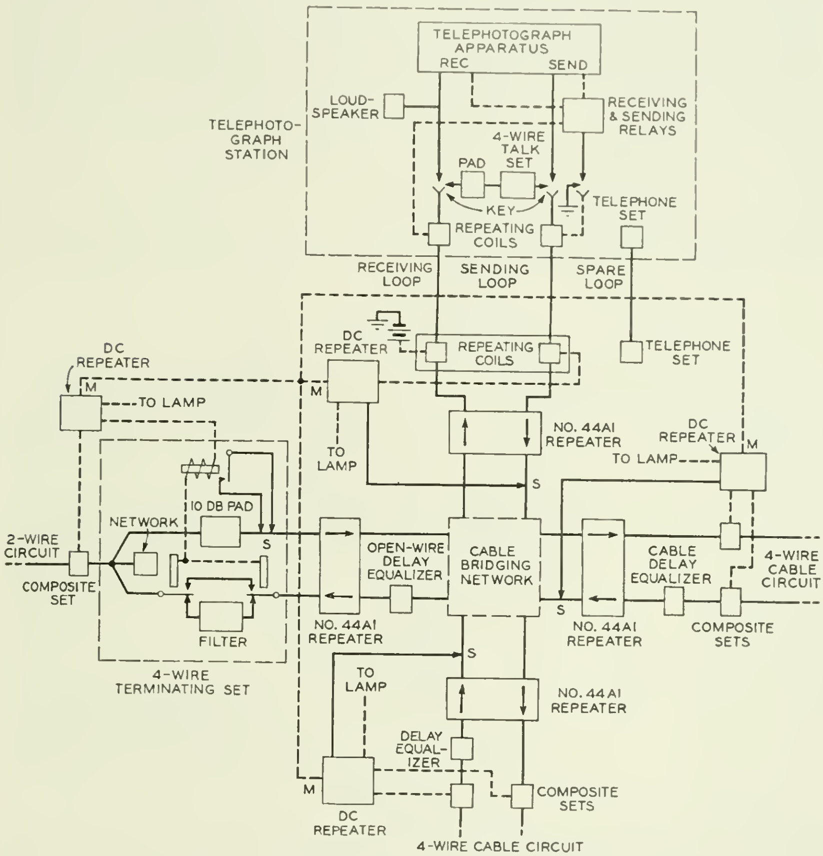


Fig. 14—Schematic diagram of arrangements at a typical bridging point.

through the use of a new form of bridging network. Although this situation could be met by employing unilateral devices such as vacuum tube amplifiers, it was found that comparable results could be obtained simply and at less expense with interconnected resistance type pads. Two designs of such networks, essentially alike except for the values of attenuation provided, are in use, one for cable and the other for open-wire circuits. These bridges are used not only at junctions of

circuits forming the network but also at all points at which telephotograph stations are connected.

A single line schematic of the type of bridging network employed is shown in Fig. 15 (upper left), and a more complete representation of a portion of the network used on cable circuits is shown in Fig. 15 (lower right). Current entering the bridge, for example at the West input, traverses three direct paths of equal attenuation and leaves at East output and branch A and branch B outputs. There are, of course, numerous indirect paths between the West input and each of the bridge

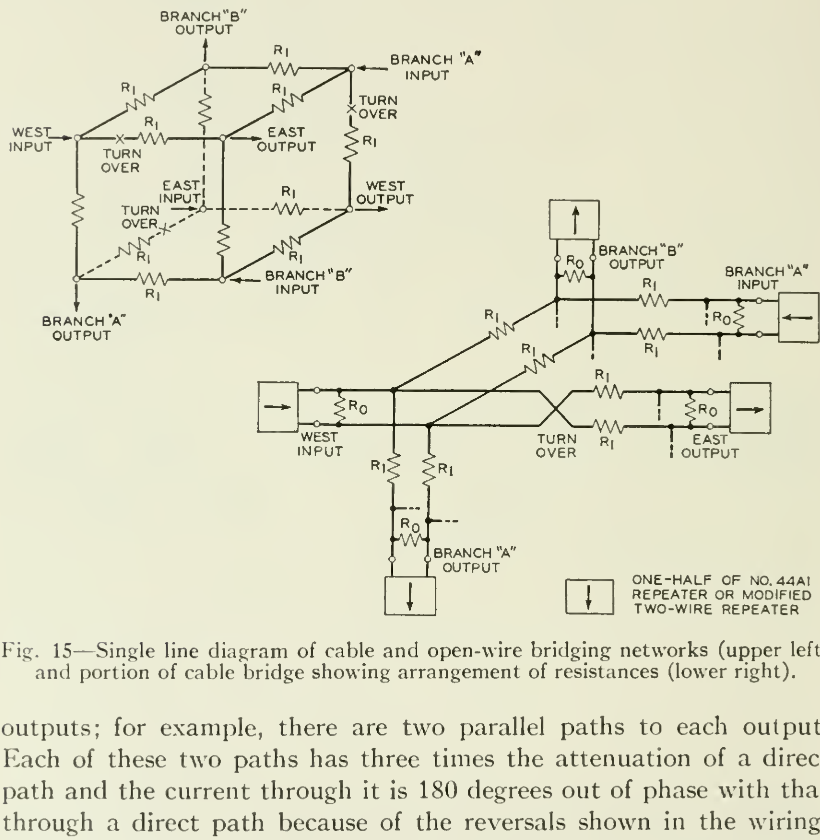


Fig. 15—Single line diagram of cable and open-wire bridging networks (upper left) and portion of cable bridge showing arrangement of resistances (lower right).

outputs; for example, there are two parallel paths to each output. Each of these two paths has three times the attenuation of a direct path and the current through it is 180 degrees out of phase with that through a direct path because of the reversals shown in the wiring. However, the aggregate of all of the indirect paths does not appreciably alter the loss between input and output of this bridge as calculated by neglecting them. It may be noted that the two directions of transmission for the same circuit are connected by six parallel paths each of which has three times the attenuation of a direct path between an input and output of the bridge. The currents through three

of these paths are 180 degrees out of phase with the currents in the others, and hence would result in infinite attenuation of the echo were it not for small unbalance currents. Measured crosstalk losses for the echo paths in excess of 70 decibels have been obtained for these bridging networks manufactured with ordinary tolerances.

Certain auxiliary features may also be incorporated in telephotograph networks to assist in their operation and perform other related functions. For example, telephotograph methods are not efficient in their present form for the rapid exchange of operating instructions; therefore telephone facilities may be associated with a telephotograph network for use by the customer in coordinating the operation of this system. Arrangements may be used whereby such voice communication may be carried on over the telephotograph circuit between picture transmissions, and loud speakers may be bridged on the circuit for monitoring purposes.

Acknowledgment

The attainment of this objective in telephotograph development and the establishment of the present leased wire network has engaged the initiative and resourcefulness of several score of individuals at the Bell Telephone Laboratories, Inc., the Western Electric Company, and the American Telephone and Telegraph Company. In reviewing the advances which have been made, the practical limitation of space has made it impossible to discuss in greater detail the various phases of the work and to render individual recognition to all who have contributed to the solution of the problems involved. Among those most intimately concerned and through whose efforts the many details have been worked out and correlated are W. A. Phelps and P. Mertz of the Bell Telephone Laboratories, Inc., and I. E. Lattimer of the American Telephone and Telegraph Company.

References

1. "The Transmission of Pictures Over Telephone Lines," H. E. Ives, J. W. Horton, R. D. Parker, and A. B. Clark, *Bell System Technical Journal*, v. 4, April 1925, pp. 187–214.
2. "Synchronization of Television," H. M. Stoller and E. R. Morton, *Bell System Technical Journal*, v. 6, Oct. 1927, pp. 604–15.
3. "Synchronization System for Two-Way Television," H. M. Stoller, *Bell System Technical Journal*, v. 9, July 1930, pp. 470–6, and discussion, p. 477.
4. "A Permanent Magnet Light Valve," G. E. Perreault, *Bell Laboratories Record*, v. 10, Aug. 1932, pp. 412–16.
5. "Certain Topics in Telegraph Transmission Theory," H. Nyquist, *A. I. E. E. Transactions*, v. 47, April 1928, pp. 617–44.
6. "A Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television," Pierre Mertz and Frank Gray, *Bell System Technical Journal*, v. 13, July 1934, pp. 464–515.
7. "An Oscillograph for Ten Thousand Cycles," A. M. Curtis, *Bell System Technical Journal*, v. 12, Jan. 1933, pp. 76–90.

8. "The Use of the Triode Valve in Maintaining the Vibration of a Tuning Fork," W. H. Eccles, *Physical Society (London) Proceedings*, v. 31, 1918–19, pp. 269.
9. "Frequency Measurement in Electrical Communication," J. W. Horton, N. H. Ricker, and W. A. Marrison, *A. I. E. E. Transactions*, v. 42, June 1923, pp. 730–41.
10. "Precision Determination of Frequency," J. W. Horton and W. A. Marrison, *I.R.E. Proceedings*, v. 16, Feb. 1928, pp. 137–54.
11. "Thermostat Design for Frequency Standards," W. A. Marrison, *I. R. E. Proceedings*, v. 16, July 1928, pp. 976–80.
12. "Long Distance Telephone Circuits in Cable," A. B. Clark and H. S. Osborne, *Bell System Technical Journal*, v. 11, Oct. 1932, pp. 520–45.
13. "The Transmission Characteristics of Open Wire Telephone Lines," E. I. Green, *Bell System Technical Journal*, v. 9, Oct. 1930, pp. 730–59.
14. "Phase Distortion and Phase Distortion Correction," Sallie P. Mead, *Bell System Technical Journal*, v. 7, April 1928, pp. 195–224.

Equivalent Networks of Negative-Grid Vacuum Tubes at Ultra-High Frequencies

By F. B. LLEWELLYN

It is shown that the equivalent network of negative-grid vacuum tubes both at low and at very high frequencies may be expressed in many different forms. Several are suggested and the advantages of two are described in some detail. One of these is closely analogous to that which is in general use at low frequencies and requires only the addition of resistive components in series both with the cathode-grid and the grid-plate capacitances to make it applicable to frequencies where transit time effects are appreciable though moderately small. The resistance in series with the grid-plate capacitance is negative in sign. In this form of the equivalent network, electron transit times do not introduce a phase angle into the amplification factor.

The paper is divided into two parts. The first gives a descriptive interpretation of the results while the second contains the mathematical manipulations.

Part I

WHEN the equivalent network of a vacuum tube is mentioned, it brings to the mind of practically every radio engineer a certain combination of resistances and capacitances together with an internal μ-generator which has become familiar through years of use. Historically, this equivalent network did not spring into being full grown like Athena from the forehead of Zeus, but was the result of a slow and painful development. The beginnings of the equivalent network of negative-grid vacuum tubes are to be found in the work of Nichols where it was pointed out that a non-linear resistance is the equivalent of a fixed resistance in series with a generator. As a second step, Van der Bijl's relation states that the plate current in a vacuum tube is a function of the plate voltage plus a constant times the grid voltage. This constant was identified with our well-known amplification factor μ and it was an easy step thereafter to combine the Van der Bijl and Nichols relations and represent the vacuum tube by the equivalent network shown in Fig. 1.

Here the cathode is located at C and the plate at P. Between them the vacuum tube is represented by the internal plate resistance r_p acting in series with the fictitious generator $\mu_0 V_g$. This equivalent naturally represents conditions between the cathode and plate at very low frequencies only, because the low-frequency impedance between the grid element and the other electrodes is so high that it can safely be disregarded. Such an equivalent network was satisfactory only so

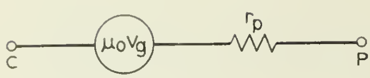


FIGURE 1

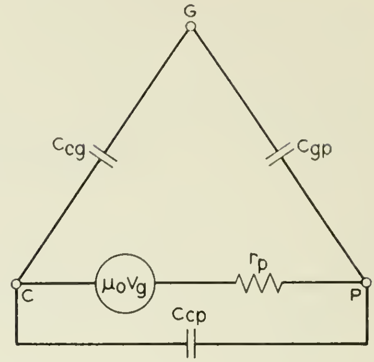


FIGURE 2

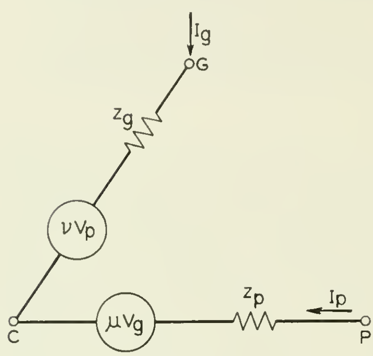


FIGURE 3

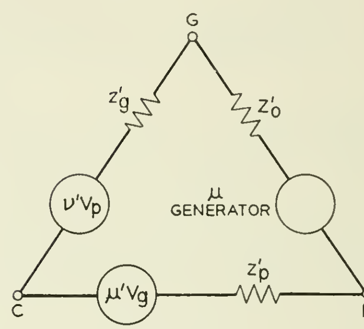


FIGURE 4

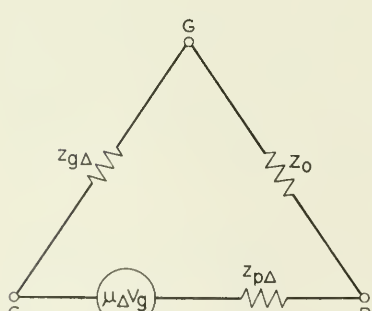


FIGURE 5

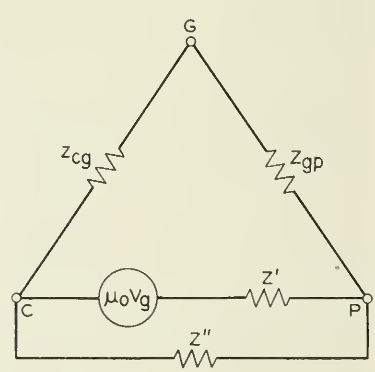


FIGURE 6

Figs. 1–6—Equivalent vacuum tube networks.

long as frequencies were low enough to allow this last approximation to remain valid. With the advent of higher frequencies it became evident that the internal tube capacitances played an important role in the operation of the device. The lengths to which early workers went to include the capacitance effects are illustrated by the complicated formulas on page 207 of Van der Bijl's well-known book on Thermionic Vacuum Tubes. Further study, however, showed that the complication could be overcome largely by a modification of the simple network of Fig. 1 so that capacitances are introduced between all three elements of the vacuum tube. The result is Fig. 2 which has been adequate in the past for all purposes. In comparatively recent years, however, increasing frequencies demand that a further revision be made.

The necessity for revision first became evident with the discovery that the impedance measured between grid and cathode when a very large condenser was placed between plate and cathode, showed an important resistive component at very high frequencies so that the simple combination of Fig. 2 involving only capacitances for the grid-cathode and grid-plate impedance was no longer valid. The tools for effecting the modification of Fig. 2 are available [1] and already have been employed to a certain extent. These tools are the result of a theoretical analysis of the motions of electrons within vacuum tubes and started from fundamentals. With the reservation that they apply strictly to planar rather than cylindrical tube structures, the results should therefore require little further modification for some time to come.

The first result of theoretical analysis was to produce an equivalent network which, on the face of it, resembles Fig. 2 only remotely, but which can be shown [1] to be exactly equivalent at low frequencies. This generalized theoretical network is shown in Fig. 3. It may be seen to consist of two branches only, which exist respectively between cathode and grid and between cathode and plate. Both branches contain internal generators and, in general, the impedance in neither branch is a pure resistance but depends upon a number of factors including the time required by electrons in traversing the vacuum tube. The immediate query which results from inspection of Fig. 3 is "What has become of the grid-plate path?" The answer to this lies in the definition of current in Fig. 3 so that the cathode-plate path is included in the network as shown. This definition of current is merely the generalized one adopted years ago by Maxwell when he

[1] F. B. Llewellyn, "Operation of Ultra-High-Frequency Vacuum Tubes," *Bell Sys. Tech. Jour.*, Vol. XIV, pp. 632–665, October 1935.

realized that a change in electric intensity produces precisely the same effect in a circuit as does an actual motion of charge. In Fig. 3 this means that the current entering the branch between cathode and grid for example consists not only of ordinary conduction current but also of displacement current so that the current in the cathode-grid mesh is the whole current flowing into the grid element. Likewise, the current in the cathode-plate mesh is the whole current flowing into the plate element of the tube.

In Part II straightforward transformation of the equations representing Fig. 3 shows that it may be represented just as well by an infinite number of other equivalent networks. Naturally our aim is to choose the form of network which is easily adaptable to the greatest number of practical applications, and the one that suggests itself primarily for this purpose contains the fewest number of internal generators. A second consideration in the choice of the best equivalent network is that the network should resemble the familiar delta equivalent of Fig. 2 as closely as may be, so that results based on that figure may be interpreted readily in terms of the more general network.

Fig. 2 is actually a modified form of a delta network. The most general delta would be the one shown in Fig. 4 which consists of three series branches, each containing an internal generator in series with an impedance. When the mathematical transformations from Fig. 3 to Fig. 4 are carried through, it is found that a proper choice of definitions for the various impedances reduces Fig. 4 to the network shown in Fig. 5. Here only one internal generator remains, but that generator acts in series with the internal plate impedance of the tube so that Fig. 5 does not quite conform to the popular network where a capacitance is assumed to shunt the internal generator by acting directly between plate and cathode. However, again it can be shown that Fig. 5 may be transformed to Fig. 6 and by a proper choice of the two impedances Z' and Z'', the internal generator reduces merely to our familiar low-frequency amplification factor multiplied by the grid potential variation.

Thus Fig. 6 with the associated definitions of impedance represents the generalized form of the equivalent network of negative-grid vacuum tubes and is valid until the velocity of the electrons approaches that of light or until the distance between elements of the vacuum tube becomes comparable to the free-space wave-length of any ultra-high frequency considered. The expressions for the various impedances in Fig. 6 are naturally long and complicated. However, at frequencies where the effects of transit time of the electrons are only moderately important, the complication reduces enormously and we have Fig. 7.

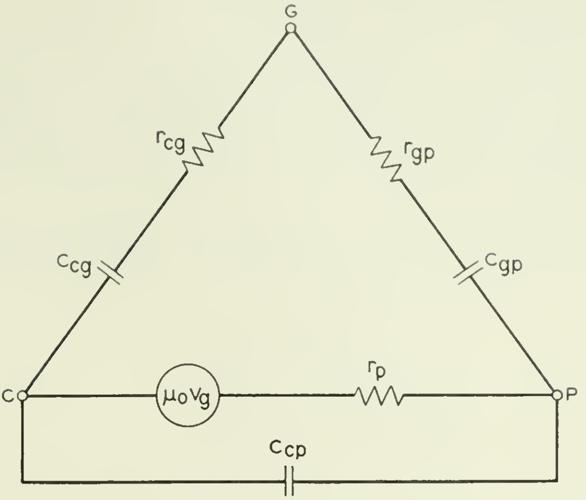


Fig. 7—General equivalent of vacuum tubes valid for moderately high frequencies.

This is nearly identical with the well-known equivalent which was shown in Fig. 2, and the modification consists only of the addition of series resistors in the internal cathode-grid and cathode-plate paths. A phase angle in the amplification factor is avoided by the resistances in series with the capacitances.[2] The impedance in series with the μ-generator is the well-known internal plate resistance as given by the slope of the static characteristic of the tube. The capacitances are likewise those we have used all along at lower frequencies, but the mathematics now enables their dielectric constants to be computed. Fig. 7 will be found to be valid at any of the frequencies for which negative-grid tubes are now contemplated for commercial application, including those where the transit angle is about half a radian.

Much has been said and written of late years about the active grid loss.[1, 3, 4, 5] In Fig. 7 this would be determined by placing a large

[2] It can be shown that measurements published in a paper "Phase Angle of Vacuum Tube Transconductance," F. B. Llewellyn, *Proc. I. R. E.*, Vol. 22, August 1934, may be interpreted as well in terms of the phase angle of the grid-plate impedance. If the phase angle of the latter is α, and the angle measured for the paper is θ, then

$$\sin^2 \alpha = 1 - \frac{C'}{C} + \frac{\sin \phi \cos \phi \tan \theta}{\omega^2 LC},$$

where C' is the grid-plate capacitance of the cold tube, C of the hot tube, and ϕ is the phase angle of the inductive branch of the tuned circuit used in the experiments.

[1] Loc. cit.

[3] J. G. Chaffee, "The Determination of Dielectric Properties at Very High Frequencies," *Proc. I. R. E.*, Vol. 22, August 1934.

[4] W. R. Ferris, "Input Resistance of Vacuum Tubes as Ultra-High Frequency Amplifiers," *Proc. I. R. E.*, Vol. 24, January 1936.

[5] D. O. North, "Analysis of the Effects of Space Charge on Grid Impedance," *Proc. I. R. E.*, Vol. 24, January 1936.

condenser between cathode and plate and measuring the input impedance. The result of computing the impedance from Fig. 7 agrees with that formerly presented [1] as, of course, it should, since both results were derived from the same fundamental analysis. This agreement, however, is mentioned by way of giving an example which checks the algebraic manipulations which were employed in arriving at Fig. 7.

Finally the values of the various elements in Fig. 7 are summarized in Table I. The formulas naturally are very long and their greatest

TABLE I

Referring to Fig. 7

Let C_{cg}', C_{gp}', C_{cp}' be capacitances of cold tube,

$$y = \frac{x_p}{x_c} \text{ be ratio of } g - p \text{ to } c - g \text{ spacing,}$$

$$h = \frac{T_p}{T_c} \text{ be ratio of } g - p \text{ to } c - g \text{ transit time,}$$

$$M = \frac{4}{3}(y - h^3)(1 + h) - 2h^2 + h^4,$$

$$N = (y - h^3)(9 + 44h + 45h^2) - 51h^2 - 105h^3 - 27h^4 + 27h^5,$$

$$C_{cp} = \frac{4}{3} C_{cp}' \left[\frac{1 + y + \dfrac{\mu_0 y}{y - h^3}}{1 + M + \mu_0} \right] \left[1 + h - \frac{3}{2} \frac{h^2}{y - h^3} \right],$$

$$r_p = \text{same as at low frequencies,}$$

$$C_{cg} = C_{cg}' \left[\frac{1 + y + \dfrac{\mu_0 y}{y - h^3}}{1 + M + \mu_0} \right] \frac{M}{y},$$

$$r_{cg} = \left[\frac{r_p(y - h^3)}{45\mu_0(1 + M + \mu_0)M^2} \right] \left[(\mu_0 + 1)N + \frac{45\mu_0 h^4}{(y - h^3)} M \right],$$

$$C_{gp} = C_{gp}' \left[\frac{1 + y + \dfrac{\mu_0 y}{y - h^3}}{1 + M + \mu_0} \right],$$

$$r_{gp} = - \left[\frac{r_p(y - h^3)}{45\mu_0(1 + M + \mu_0)} \right] \left[N - \frac{45\mu_0 h^4}{y - h^3} \right].$$

use probably is in describing the simple circuit of Fig. 7 where the values of the various elements can actually be measured or computed as convenient.

The easiest way to visualize the equations is to apply them to a special case which can be approached experimentally; namely the condition that the time required by electrons in moving from grid to plate is much shorter than the cathode-grid time. When this is the case, the formulas reduce to those shown in Table II. These show

TABLE II

REFERRING TO FIG. 7

Let C_{cg}', C_{gp}', C_{cp}' be capacitances of cold tube,

$$y = \frac{x_p}{x_c} = \text{ratio of } g - p \text{ to } c - g \text{ spacing},$$

$$h = \frac{T_p}{T_c} = \text{ratio of } g - p \text{ to } c - g \text{ transit time}.$$

Then when $h \rightarrow 0$:

$$C_{cp} = \frac{4}{3} C_{cp}' \left[\frac{1 + y + \mu_0}{1 + \frac{4}{3} y + \mu_0} \right],$$

$$r_p = \text{same as at low frequencies},$$

$$C_{cg} = \frac{4}{3} C_{cg}' \left[\frac{1 + y + \mu_0}{1 + \frac{4}{3} y + \mu_0} \right],$$

$$r_{cg} = \frac{9}{80} \frac{r_p}{\mu_0} \left[\frac{1 + \mu_0}{1 + \frac{4}{3} y + \mu_0} \right],$$

$$C_{gp} = C_{gp}' \left[\frac{1 + y + \mu_0}{1 + \frac{4}{3} y + \mu_0} \right],$$

$$r_{gp} = -\frac{1}{5} \frac{r_p}{\mu_0} \left[\frac{y^2}{1 + \frac{4}{3} y + \mu_0} \right].$$

that the cathode-plate and cathode-grid capacitances have dielectric constants greater than unity, but that the grid-plate capacitance has a dielectric constant less than unity. The cathode-grid resistance is positive, and the grid-plate resistance is negative.

The outstanding result of this investigation of the network representing the negative-grid tube is the demonstration of the slight modification required in our conventional network to make it accurate even in the ultra-high-frequency range. The amplification factor is the familiar low-frequency one, and at moderately high frequencies, the only alteration needed in the conventional diagram is the addition of two small but very important resistances, one in the cathode-grid path and one in the grid-plate path, where the resistance in the latter path is negative in sign.

PART II

In a recent paper,[1] general equations have been derived which describe the behavior of vacuum tubes at ultra-high frequencies. In

[1] Loc. cit.

the case of negative-grid triodes and referred to Fig. 8, these equations take the general form:

$$V_p + \mu V_g = I_p z_p, \tag{1}$$

$$V_g - \nu V_p = I_g z_g, \tag{2}$$

where

$$\left.\begin{aligned} \mu &= \frac{(Z_1 + Z_2) - (Z_2 + Z_3 + Z_c)}{Z_c + Z_g}, \\ z_p &= \frac{(Z_2 + Z_3 + Z_c)Z_g + (Z_1 + Z_2)Z_c}{Z_c + Z_g}, \\ \nu &= \frac{Z_c}{Z_2 + Z_3 + Z_c}, \\ z_g &= \frac{(Z_2 + Z_3 + Z_c)Z_g + (Z_1 + Z_2)Z_c}{Z_2 + Z_3 + Z_c}, \end{aligned}\right\} \tag{3}$$

and the Z's may be expressed in terms of the tube geometry and d-c. current or voltage by means of equations (80)–(84) in the reference. In these relations the currents, I_p and I_g, denote the total current reaching plate or grid, respectively, and hence include both the conduction current carried by the electrons themselves and the displacement current arising from the change of electric force. With this meaning of current, (1) and (2) contain the complete description of the performance of the tube, and separate consideration of the grid-plate current, usual in low-frequency methods, is unnecessary because that current is already included in I_p in (1).

The equivalent network represented by (1) and (2) is shown in Fig. 8. Only two currents are involved, I_p and I_g, but, also two internal generators, μV_g and νV_p, are required. For some purposes, an equivalent network which corresponds more nearly with the usual low-frequency delta arrangement is desirable. Such an equivalent may be obtained from (1) and (2) in conjunction with Fig. 9 which shows the relation between currents in a delta network and those of Fig. 8. In Fig. 9 no restriction is yet placed upon the three currents, I_1, I_2 and I_3, so that in general they all may be allowed to include both conduction and displacement components. From Fig. 9

$$I_p = I_1 + I_2, \tag{4}$$

$$I_g = I_3 - I_2. \tag{5}$$

Here are two equations expressing the three unknowns, I_1, I_2 and I_3, in terms of the currents I_p and I_g, which are assumed to be known.

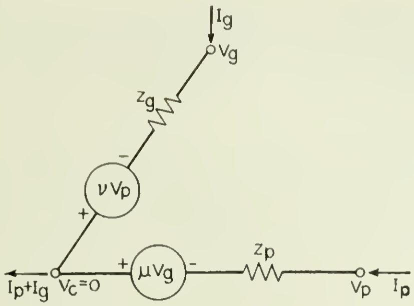


FIGURE 8

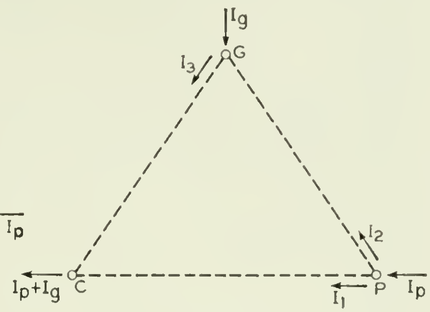


FIGURE 9

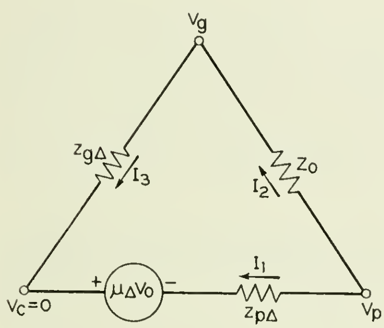


FIGURE 10

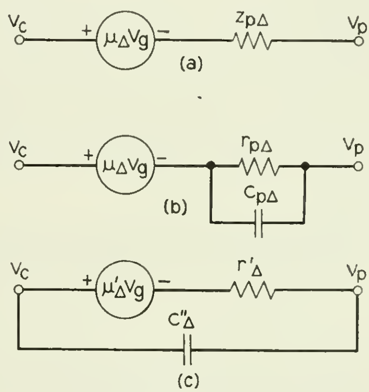


FIGURE 11

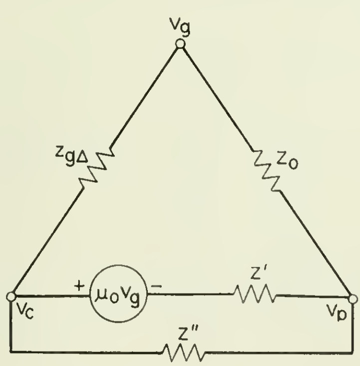


FIGURE 12

Fig. 8—Equivalent network of vacuum tube from equations (1) and (2).
Fig. 9—Relation between currents of delta network and those of Fig. 8.
Fig. 10—Equivalent delta network.
Fig. 11—Equivalent cathode-plate paths.
Fig. 12—Modified delta network equivalent to Fig. 10.

It is obvious that a third equation is needed before the unknowns can be found. But (4) and (5) express all the relationships that are necessary for the equivalence of Figs. 8 and 9. It follows that we are at liberty to impose arbitrarily a third restriction upon the currents in Fig. 9.

The choice of this restriction may be made in many ways, each resulting in a different network, all equivalent however to Fig. 8 and to the vacuum tube. For example, I_1 might be defined as consisting of conduction current only. Such a choice might seem at first sight to be a desirable one because it corresponds rather well with the conception of the cathode-plate path as being determined by electron movement at low frequencies. If it were adopted, however, the generalized network resulting would be found to be quite awkward, involving two or more internal generators and complex amplification factors.

The simplest network would be the one involving the fewest number of internal generators, and the restriction adopted in the following analysis for the currents in Fig. 9 is made with that object in view. The result, as will be shown, corresponds at low frequencies with the usual concept of the tube, and gives a high-frequency network where neither the cathode-grid nor the grid-plate paths contain internal generators.

The restriction which accomplishes this result is obtained by placing

$$V_p - V_g = I_2 Z_0, \tag{6}$$

so that (4), (5) and (6) determine the internal currents, I_1, I_2 and I_3, in terms of the external currents I_p and I_g, and the, as yet, arbitrary impedance Z_0.

The solution of (1) to (6) yields

$$V_p + V_g \left[\frac{\mu + \dfrac{z_p}{Z_0}}{1 - \dfrac{z_p}{Z_0}} \right] = I_1 \left[\frac{z_p}{1 - \dfrac{z_p}{Z_0}} \right], \tag{7}$$

$$V_g - V_p \left[\frac{\nu - \dfrac{z_g}{Z_0}}{1 - \dfrac{z_g}{Z_0}} \right] = I_3 \left[\frac{z_g}{1 - \dfrac{z_g}{Z_0}} \right]. \tag{8}$$

The choice of (6) eliminates the internal generator from the grid-plate path, but still leaves the impedance Z_0 to be defined at will. From (8)

it is evident that the internal generator is eliminated from the cathode-grid path if Z_0 is chosen so that

$$\nu = z_g/Z_0. \tag{9}$$

The impedance Z_0 will accordingly be taken to be defined by (9). The result is that the fundamental equations for Fig. 9 become:

$$V_p + \mu_\Delta V_g = I_1 z_{p\Delta}, \tag{10}$$

$$V_g = I_3 z_{g\Delta}, \tag{11}$$

$$V_p - V_g = I_2 Z_0, \tag{12}$$

where

$$\left.\begin{aligned} \mu_\Delta &= \frac{Z_1 - Z_3}{Z_g}, \\ z_{p\Delta} &= Z_2 + Z_3 + Z_c + \frac{Z_c}{Z_g}(Z_1 + Z_2), \\ z_{g\Delta} &= Z_g + \frac{Z_c(Z_1 + Z_2 + Z_g)}{Z_2 + Z_3}, \\ Z_0 &= Z_1 + Z_2 + \frac{Z_g}{Z_c}(Z_2 + Z_3 + Z_c). \end{aligned}\right\} \tag{13}$$

The various definitions in (13) are seen to be slightly simpler than those in (3) but it must be remembered that the delta-network involves one more current-path than the original ultra-high-frequency network, Fig. 1. The delta corresponding to (10)–(13) is shown in Fig. 10.

At frequencies only moderately high, all of the impedances in Fig. 10 are composed of combinations of ordinary resistances and capacitances. Both $z_{g\Delta}$ and Z_0 consist of a condenser and resistor in series.

In the case of the cathode-plate path in Fig. 10, the equivalent combination of resistance and capacitance may be represented by either a parallel combination of resistance and capacitance which is in series with the μ-generator, as shown at (*b*) in Fig. 11, or a μ-generator of different value acting in series with a resistance, and the whole being shunted by a capacitance connected between cathode and plate, as shown at (*c*) in Fig. 11. The latter picture is in more strict accord with conventional practice but the mathematical relationships involve a choice of definitions for μ. It is found by trial that this choice may be made so that μ in Fig. 12 is defined merely as μ_0, its low-frequency value, and is independent of frequency. When this definition is adopted, we have in general the equivalent network of Fig. 12 which holds at high as well as low frequencies. This differs from Fig. 10 in the cathode-plate path only, and Z', Z'' are defined as follows, where μ_0

is the low-frequency amplification factor:

$$Z' = \mu_0 \left[\frac{(Z_2 + Z_3 + Z_c)Z_g + (Z_1 + Z_2)Z_c}{Z_1 - Z_3} \right], \tag{14}$$

$$Z'' = \mu_0 \left[\frac{(Z_2 + Z_3 + Z_c)Z_g + (Z_1 + Z_2)Z_c}{\mu_0 Z_g - Z_1 + Z_3} \right]. \tag{15}$$

Figure 7 is the form taken by Fig. 12 for moderately high frequencies where transit angle effects are just beginning to become noticeable.

In general, the formulas for the impedances in the delta network are just as long as in the original network of Fig. 8. The delta may, however, have an advantage in view of its wide use in low-frequency work, and of the fact that the amplification factor for the delta can be expressed without involving the transit angle, and hence does not contain a phase shift.

Forces of Oblique Winds on Telephone Wires

By J. A. CARR

In aerial line design it is advantageous to know the effect of oblique winds as well as cross winds. This paper gives the results of wind tunnel tests made on 0.104-inch and 0.165-inch diameter wires for each 10° angle of obliquity between 0° and 90° using wind velocities of 30 to 90 miles per hour in steps of 10 miles per hour. These results are then analyzed to determine (1) their compliance with the law of dynamic similarity and (2) the magnitudes of the various wind components. From these analyzed results an expression is developed for the force of oblique winds in terms of the component normal to the wires.

IN connection with studies of wire arrangements on open-wire lines [1] which Bell Telephone Laboratories have had under way for some time, it became necessary to evaluate the resistance of wires to winds. The method of evaluating the force of winds normal to the wires has been studied by many investigators and there is a considerable amount of data in the literature on this subject. The contrary was found to be true in the case of oblique winds or those not normal to the line. This latter case has been described briefly in the records of a test made in the National Physical Laboratories [2] (British) on a 0.375-inch diameter smooth wire at a wind velocity of 40 feet per second (27.3 m.p.h.) with the wire at angles to the wind ranging from 0° to 90° (normal) in steps of 10° and also in the records of M. Gustave Eiffel,[3] who made a similar test at somewhat higher velocities. Since the wires we are concerned with range from about 0.1 to 0.2 of an inch in diameter and the wind velocity ranges from about 30 to 90 miles per hour, it appeared desirable to conduct a series of wind tunnel tests that would extend these data and more fully meet our requirements. Tests along these lines were arranged with the Guggenheim School of Aeronautics at New York University.[4] Subsequently, a series of tests was made in the New York University wind tunnel on 0.104-inch and 0.165-inch diameter smooth copper wires for each 10° angle ranging from 0° to 90° using wind velocities of 30 to 90 miles per hour in steps

[1] "Motion of Telephone Wires in Wind," D. A. Quarles, *Bell System Technical Journal*, April 1930.
[2] Reports and Memoranda No. 307, January 1917, entitled "Tests on Smooth and Stranded Wires Inclined to the Wind Direction," by E. F. Relf and C. H. Powell.
[3] "Nouvelles Recherches sur La Résistance De L'Air et L'Aviation," book by M. G. Eiffel.
[4] These tests were conducted by Professor Alexander Klemin and his associates.

of 10 miles per hour. The readings so obtained are here studied and analyzed.

The wire set-up used in the wind tunnel is shown in the accompanying picture (Fig. 1) and is similar to that used by Eiffel.[3] It comprised a five-foot frame of 0.375-inch diameter steel in which five wires of either the 0.104-inch or 0.165-inch size were mounted. A spacing of 2.75 inches was used between the centers of the wires. The frame of wires

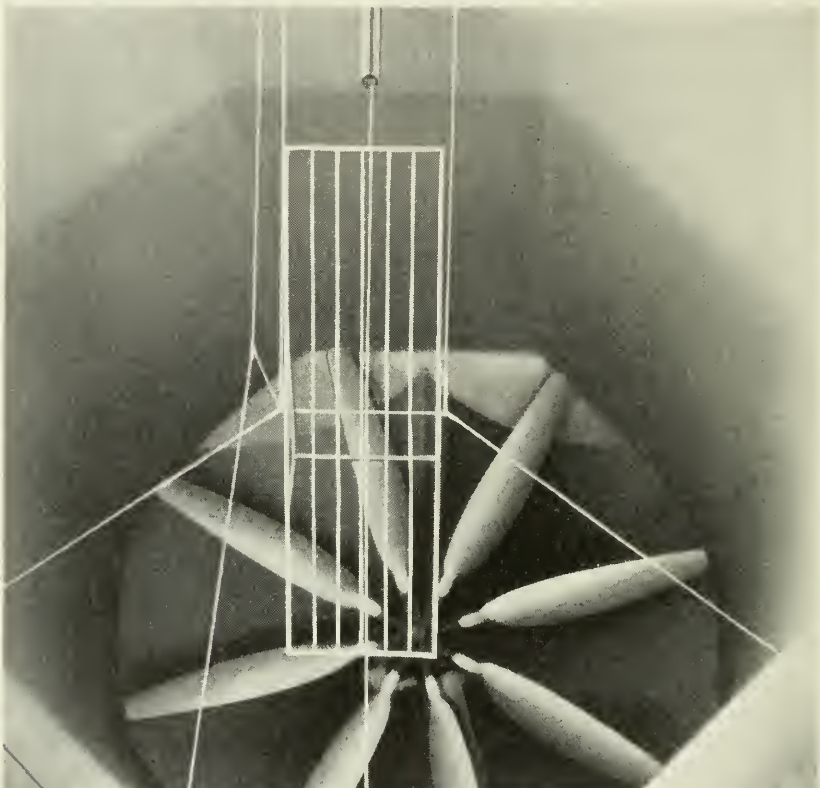


Fig. 1—Wind tunnel setup.

was installed in the approximate center of the nine-foot section of the tunnel with the shorter axis of the frame in a horizontal plane and perpendicular to the axis of the tunnel. This arrangement was convenient for connections to the weighing mechanism and permitted the frame to be rotated about its short axis.

During each step in the test the horizontal (drag) and vertical (lift) forces on the frame of wires were measured at least three times. At the

completion of the series of tests on each size of wire the test lengths were cut out of the frame leaving about 2 inches of each wire at each end and the series of tests repeated so that the net drag and net lift figures could be determined. The purpose of leaving the short lengths of wire at each end was to provide a correction for the interference effects introduced at the ends of the wires. This practice was probably effective as it will be shown later that, where comparisons could be made, the results obtained in these tests agree satisfactorily

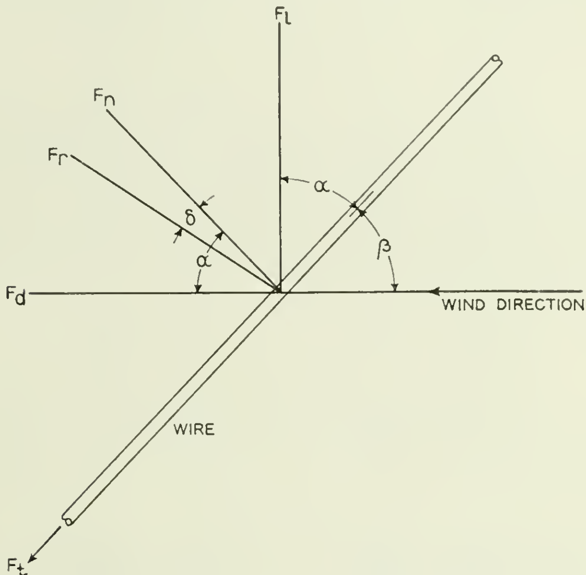


Fig. 2—Force components of wind on wires.

F_d = Force along the direction of wind—drag.
F_l = Force across the direction of wind—lift.
F_r = Resultant force.
F_n = Force normal to the wire.
F_t = Force along the wire—tangential.

with those previously obtained. This is true even though the frame comprised a fairly large proportion of the total resistance especially when the angle between the wind and the frame was small. The case when the wind and wires were parallel (tangential) has been omitted from the results because of the large proportion of the total resistance as well as the interference offered by the frame and because of its relative unimportance to this study.

Figure 2 shows the forces on the wires to which consideration has

been given here. From this diagram it can be seen that with values of the net drag (F_d) and lift (F_l), the normal force (F_n), the resultant force (F_r), the tangential force (F_t) and the angle (δ) between the resultant and normal forces can be determined through the following relationships:

$$F_n = F_l \sin \alpha + F_d \cos \alpha = F_r \cos \delta, \tag{1}$$

$$F_r = \sqrt{F_d^2 + F_l^2} = \sqrt{F_n^2 + F_t^2}, \tag{2}$$

$$F_t = F_d \sin \alpha - F_l \cos \alpha = F_r \sin \delta, \tag{3}$$

$$\delta = \alpha - \text{arc tan}\,\frac{F_l}{F_d} = \text{arc tan}\,\frac{F_t}{F_n}\cdot \tag{4}$$

The only term in these equations which is not defined above is α. This is the angle between the wire and the normal to the wind or between the wind and the normal to the wire.

The data determined through the use of these relationships are given in the accompanying Tables I (0.104-inch wire) and II (0.165-inch wire). The forces in these tables are given in terms of pounds per foot of wire.

TABLE I

0.104-Inch Diameter Wire

V	α	F_n	F_t	F_r	δ	$\frac{F_n}{(\cos \alpha)^2}$	K
30	0	0.0194	0.000000	0.0194	0	0.0194	0.000207
	10	0.0189	0.000431	0.0190	1.3	0.0195	0.000208
	20	0.0177	0.001203	0.0178	3.9	0.0200	0.000214
	30	0.0149	0.001940	0.0151	7.3	0.0198	0.000212
	40	0.0119	0.002540	0.0122	12.0	0.0203	0.000217
	50	0.0083	0.002650	0.0086	18.0	0.0200	0.000215
	60	0.0047	0.002230	0.0052	25.4	0.0188	0.000201 $\bar{X}=0.000211$
	70	0.0021	0.001560	0.0026	36.7	0.0180	0.000194
	80	0.0008	0.001050	0.0013	53.9	0.0265	0.000283 $\bar{X}=0.000217$
	90	—	—	—	90.0	—	—
40	0	0.0358	0.000000	0.0358	0	0.0358	0.000215
	10	0.0334	0.000570	0.0334	1.1	0.0345	0.000207
	20	0.0300	0.001520	0.0301	2.9	0.0341	0.000204
	30	0.0252	0.002510	0.0254	5.6	0.0336	0.000202
	40	0.0199	0.003150	0.0202	9.0	0.0340	0.000204
	50	0.0144	0.003360	0.0149	13.0	0.0350	0.000210
	60	0.0087	0.003180	0.0093	20.0	0.0348	0.000210 $\bar{X}=0.000207$
	70	0.0041	0.002450	0.0049	30.0	0.0350	0.000211
	80	0.0014	0.001450	0.0020	46.1	0.0467	0.000279 $\bar{X}=0.000216$
	90	—	—	—	90.0	—	—

TABLE 1—(*Continued*)

V	α	F_n	F_t	F_r	δ	$\frac{F_n}{(\cos\alpha)^2}$	K	
50	0	0.0578	0.000000	0.0578	0	0.0578	0.000222	
	10	0.0550	0.000863	0.0551	0.9	0.0570	0.000218	
	20	0.0494	0.001980	0.0495	2.3	0.0565	0.000215	
	30	0.0412	0.003170	0.0414	4.4	0.0550	0.000211	
	40	0.0323	0.003970	0.0325	7.0	0.0550	0.000207	
	50	0.0233	0.004350	0.0235	10.7	0.0565	0.000217	
	60	0.0147	0.004010	0.0153	15.2	0.0586	0.000226	$\bar{X}=0.000217$
	70	0.0070	0.003000	0.0076	23.2	0.0600	0.000230	
	80	0.0026	0.002120	0.0033	40.0	0.0867	0.000339	$\bar{X}=0.000232$
	90	—	—	—	90.0	—	—	
60	0	0.0854	0.000000	0.0854	0	0.0854	0.000228	
	10	0.0787	0.000950	0.0788	0.7	0.0812	0.000217	
	20	0.0708	0.002350	0.0710	1.9	0.0801	0.000214	
	30	0.0592	0.003730	0.0595	3.6	0.0790	0.000210	
	40	0.0475	0.004960	0.0476	6.0	0.0811	0.000211	
	50	0.0346	0.005340	0.0352	8.7	0.0842	0.000224	
	60	0.0221	0.005100	0.0230	12.8	0.0884	0.000236	$\bar{X}=0.000220$
	70	0.0117	0.004270	0.0125	20.0	0.1000	0.000267	
	80	0.0039	0.002710	0.0047	35.3	0.1288	0.000345	$\bar{X}=0.000239$
	90	—	—	—	90.0	—	—	
70	0	0.1183	0.000000	0.1183	0	0.1183	0.000233	
	10	0.1130	0.001180	0.1131	0.6	0.1170	0.000225	
	20	0.1020	0.002860	0.1023	1.6	0.1155	0.000227	
	30	0.0873	0.004950	0.0873	3.2	0.1164	0.000228	
	40	0.0650	0.005800	0.0652	5.1	0.1105	0.000217	
	50	0.0467	0.006070	0.0472	7.4	0.1130	0.000222	
	60	0.0298	0.005900	0.0306	11.1	0.1190	0.000234	$\bar{X}=0.000227$
	70	0.0159	0.004930	0.0168	17.0	0.1370	0.000267	
	80	0.0063	0.003730	0.0072	31.3	0.2100	0.000410	$\bar{X}=0.000251$
	90	—	—	—	90.0	—	—	
80	0	0.1570	0.000000	0.1570	0	0.1570	0.000236	
	10	0.1510	0.001310	0.1510	0.5	0.1565	0.000232	
	20	0.1390	0.003400	0.1392	1.4	0.1575	0.000237	
	30	0.1213	0.005720	0.1215	2.7	0.1617	0.000243	
	40	0.0918	0.007050	0.0920	4.4	0.1560	0.000235	
	50	0.0615	0.007100	0.0618	6.6	0.1485	0.000223	
	60	0.0385	0.006600	0.0392	9.7	0.1540	0.000231	$\bar{X}=0.000234$
	70	0.0208	0.005600	0.0216	15.0	0.1778	0.000267	
	80	0.0086	0.004500	0.0096	28.0	0.2865	0.000428	$\bar{X}=0.000259$
	90	—	—	—	90.0	—	—	
90	0	0.2010	0.000000	0.2010	0	0.2010	0.000239	
	10	0.1921	0.001670	0.1921	0.5	0.1990	0.000235	
	20	0.1788	0.003740	0.1790	1.2	0.2020	0.000240	
	30	0.1546	0.005950	0.1550	2.2	0.2060	0.000245	
	40	0.1190	0.008220	0.1192	3.9	0.2030	0.000241	
	50	0.0840	0.008500	0.0844	5.8	0.2040	0.000241	
	60	0.0514	0.007930	0.0518	8.8	0.2060	0.000168	$\bar{X}=0.000230$
	70	0.0281	0.006530	0.0288	13.1	0.2400	0.000285	
	80	0.0127	0.005500	0.0136	24.0	0.4210	0.000500	$\bar{X}=0.000266$
	90	—	—	—	90.0	—	—	

TABLE II

0.165-Inch Diameter Wire

V	α	F_n	F_t	F_r	δ	$\frac{F_n}{(\cos\alpha)^2}$	K	
30	0	0.0328	0.000000	0.0328	0	0.0328	0.000221	
	10	0.0319	0.000559	0.0319	1.0	0.0329	0.000222	
	20	0.0287	0.001153	0.0287	2.3	0.0325	0.000219	
	30	0.0243	0.001710	0.0245	4.0	0.0324	0.000218	
	40	0.0187	0.002158	0.0188	6.6	0.0319	0.000215	
	50	0.0133	0.002450	0.0136	10.4	0.0322	0.000217	
	60	0.0074	0.002540	0.0078	19.0	0.0296	0.000199	$\bar{X}=0.000216$
	70	0.0032	0.001900	0.0037	30.8	0.0276	0.000184	
	80	0.0010	0.001330	0.0018	47.7	0.0331	0.000223	$\bar{X}=0.000213$
	90	—	—	—	90.0	—	—	
40	0	0.0606	0.000000	0.0606	0	0.0606	0.000230	
	10	0.0570	0.000800	0.0570	0.8	0.0580	0.000223	
	20	0.0508	0.001600	0.0508	1.8	0.0575	0.000218	
	30	0.0427	0.002235	0.0428	3.0	0.0570	0.000216	
	40	0.0342	0.002990	0.0343	5.0	0.0582	0.000221	
	50	0.0240	0.003880	0.0243	9.2	0.0580	0.000220	
	60	0.0155	0.004150	0.0160	15.0	0.0620	0.000235	$\bar{X}=0.000223$
	70	0.0074	0.003500	0.0082	25.2	0.0637	0.000240	
	80	0.0023	0.001840	0.0029	39.2	0.0760	0.000289	$\bar{X}=0.000232$
	90	—	—	—	90.0	—	—	
50	0	0.0969	0.000000	0.0969	0	0.0969	0.000235	
	10	0.0907	0.001110	0.0908	0.7	0.0940	0.000226	
	20	0.0819	0.002145	0.0820	1.5	0.0925	0.000235	
	30	0.0692	0.003260	0.0693	2.7	0.0923	0.000224	
	40	0.0562	0.004330	0.0563	4.4	0.0956	0.000232	
	50	0.0408	0.005060	0.0410	7.1	0.0986	0.000229	
	60	0.0259	0.005490	0.0265	12.0	0.1035	0.000251	$\bar{X}=0.000233$
	70	0.0119	0.004635	0.0127	21.4	0.1025	0.000247	
	80	0.0043	0.002790	0.0049	34.0	0.1430	0.000346	$\bar{X}=0.000247$
	90	—	—	—	90.0	—	—	
60	0	0.1425	0.000000	0.1425	0	0.1425	0.000240	
	10	0.1350	0.001420	0.1350	0.6	0.1393	0.000234	
	20	0.1250	0.002900	0.1251	1.3	0.1415	0.000238	
	30	0.1059	0.004250	0.1060	2.3	0.1410	0.000238	
	40	0.0862	0.005740	0.0866	3.8	0.1470	0.000246	
	50	0.0607	0.006680	0.0610	6.3	0.1470	0.000246	
	60	0.0372	0.006750	0.0378	10.3	0.01485	0.000251	$\bar{X}=0.000242$
	70	0.0180	0.005960	0.0190	18.3	0.1550	0.000259	
	80	0.0071	0.003900	0.0081	29.2	0.2360	0.000398	$\bar{X}=0.000261$
	90	—	—	—	90.0	—	—	
70	0	0.1970	0.000000	0.1970	0	0.1970	0.000244	
	10	0.1850	0.001608	0.1850	0.5	0.1908	0.000236	
	20	0.1660	0.003260	0.1660	1.1	0.1990	0.000233	
	30	0.1420	0.004960	0.1420	2.0	0.1895	0.000233	
	40	0.1180	0.006600	0.1185	3.2	0.2010	0.000249	
	50	0.0845	0.007860	0.0850	5.3	0.2050	0.000253	
	60	0.0528	0.008270	0.0535	8.9	0.2110	0.000261	$\bar{X}=0.000244$
	70	0.0270	0.007230	0.0280	15.0	0.2310	0.000285	
	80	0.0118	0.005310	0.0129	24.3	0.3940	0.000484	$\bar{X}=0.000275$
	90	—	—	—	90.0	—	—	

TABLE II—(*Continued*)

V	α	F_n	F_t	F_r	δ	$\frac{F_n}{(\cos\alpha)^2}$	K	
80	0	0.2610	0.000000	0.2610	0	0.2610	0.000247	
	10	0.2520	0.001765	0.2520	0.4	0.2610	0.000246	
	20	0.2330	0.003560	0.2330	0.9	0.2640	0.000250	
	30	0.2030	0.005310	0.2031	1.5	0.2710	0.000256	
	40	0.1650	0.007200	0.1652	2.5	0.2807	0.000266	
	50	0.1130	0.008310	0.1136	4.2	0.2740	0.000259	
	60	0.0722	0.008850	0.0727	7.0	0.2890	0.000273	$\bar{X}=0.000257$
	70	0.0400	0.008500	0.0409	12.0	0.3450	0.000324	
	80	0.0178	0.006860	0.0190	21.2	0.5930	0.000559	$\bar{X}=0.000298$
	90	—	—	—	90.0	—	—	
90	0	0.3335	0.000000	0.3335	0	0.3335	0.000250	
	10	0.3262	0.001920	0.3263	0.3	0.3380	0.000251	
	20	0.3050	0.003670	0.3050	0.7	0.3450	0.000258	
	30	0.2720	0.005680	0.2730	1.2	0.3627	0.000271	
	40	0.2235	0.007640	0.2240	1.9	0.3815	0.000285	
	50	0.1525	0.008950	0.1530	3.3	0.3690	0.000276	
	60	0.0982	0.009690	0.0985	5.6	0.3930	0.000294	$\bar{X}=0.000269$
	70	0.0562	0.009480	0.0570	9.6	0.4840	0.000359	
	80	0.0252	0.008430	0.0266	17.6	0.8400	0.000625	$\bar{X}=0.000319$
	90	—	—	—	90.0	—	—	

In characterizing the normal force of oblique winds, curves were plotted between the normal force (F_n) and the angle (α) between the wind and the normal to the wires. Figures 3 (0.104-inch wire) and 4 (0.165-inch wire) show these curves.

In studying the significance of these curves and the underlying data, consideration was first given to the extent to which the case of normal winds ($\cos\alpha = 1$) followed the law of dynamic similarity. This law states, in effect, that for any two geometrically similar bodies moving through a fluid,

$$F = \rho V^2 f\left(\frac{VD}{\nu}\right). \tag{5}$$

Here, F is the unit force at either of two similarly situated points on the two bodies, ρ is the fluid density, V is the velocity of the body, D a linear quantity depending on the dimensions of the body and ν is the kinematic viscosity of the fluid. The function (VD/ν) is the well known Reynolds number and the key to dynamic similarity requirements in model experiments made at ordinary velocities. The principle of dynamic similarity is satisfied as long as the Reynolds number is held constant.

In Report No. 102 [5] of the National Physical Laboratories is given an

[5] Reports and Memoranda No. 102, November 1914, entitled "Discussion of the Results of Measurements of the Resistance of Wires" by E. F. Relf.

empirical curve of the resistance of smooth wires to winds normal to the wire. In preparing this curve equation (5) was rewritten in terms of the total force on a diameter length of wire, namely

$$F_c = \rho V^2 D^2 f\left(\frac{VD}{\nu}\right). \tag{5a}$$

Then $F_c/\rho V^2 D^2$ was plotted as the ordinate and $\log_{10}(VD/\nu)$ as the

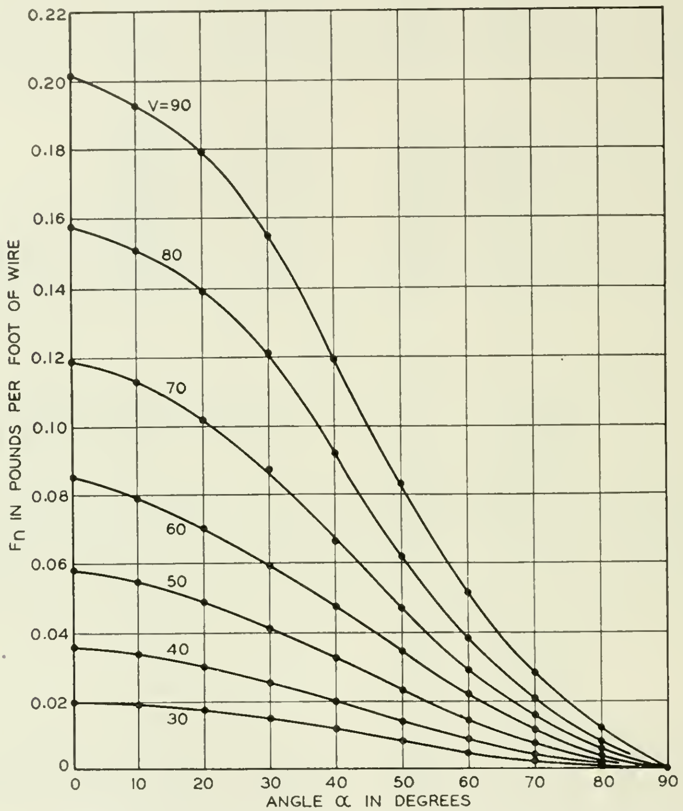


Fig. 3—Graphic relation between the force (F_n) normal to the wire and the angle (α) of wind direction from normal—0.104-inch diameter wire.

abscissa. If the normal wind data obtained in the studies of Bell Telephone Laboratories were consistent with those given in this Report No. 102 [5] a plot of the points determined through the use of equation (5a) should lie reasonably close to this curve. Figure 5 gives this curve and a plot of the normal wind results. These points appear to show satisfactory agreement.

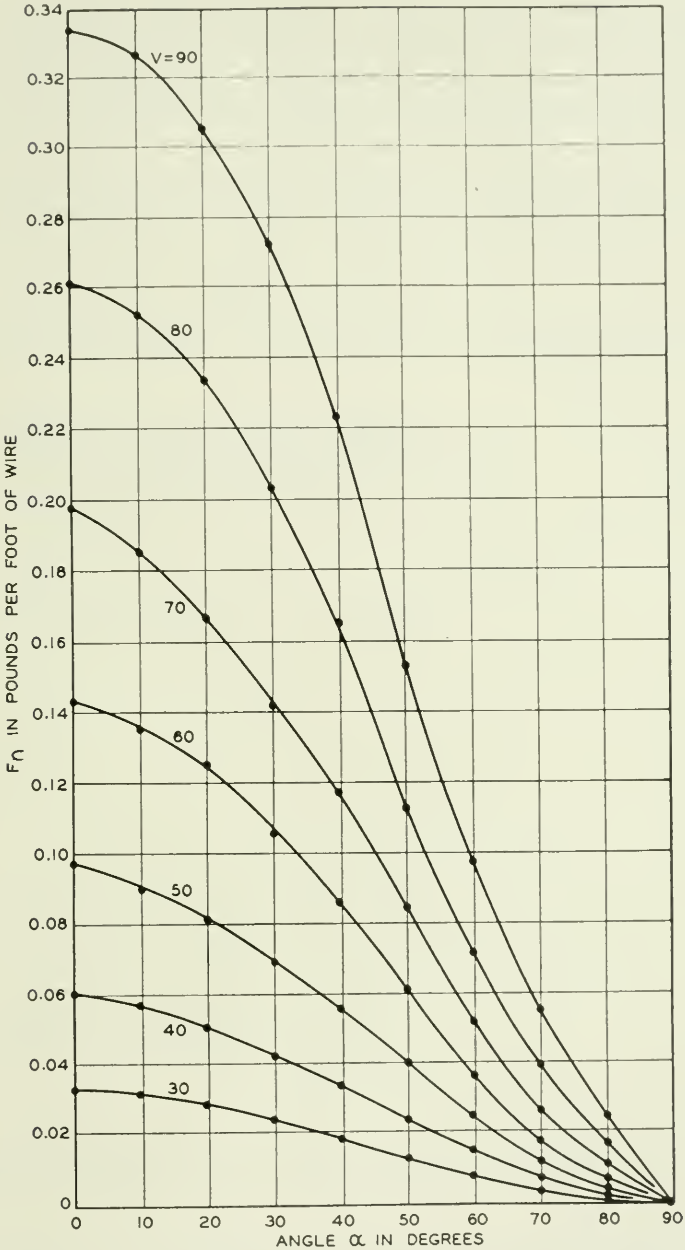


Fig. 4—Graphic relation between the force (F_n) normal to the wire and the angle (α) of wind direction from normal—0.165-inch diameter wire.

In the case of oblique winds it was found from an analysis of the data that the normal component of the force was closely proportional to $\cos^2 \alpha$ over a range of angles from 0° to 60° from normal in the case of each actual velocity and each size of wire. The values of $F_n/\cos^2 \alpha$ are given in the accompanying Tables I (0.104-inch wire) and II (0.165-inch wire). This agrees with the results obtained by Relf and Powell.[2] However, they used only one size of wire and one wind velocity in their tests. Expressing this result in terms of the normal force gives,

$$F = (K' \cos^2 \alpha)_{V,\ D=\text{constant}}.$$

This empirical expression suggested that a form of relation existed similar to that for the case of normal winds (equation 5a). Studying

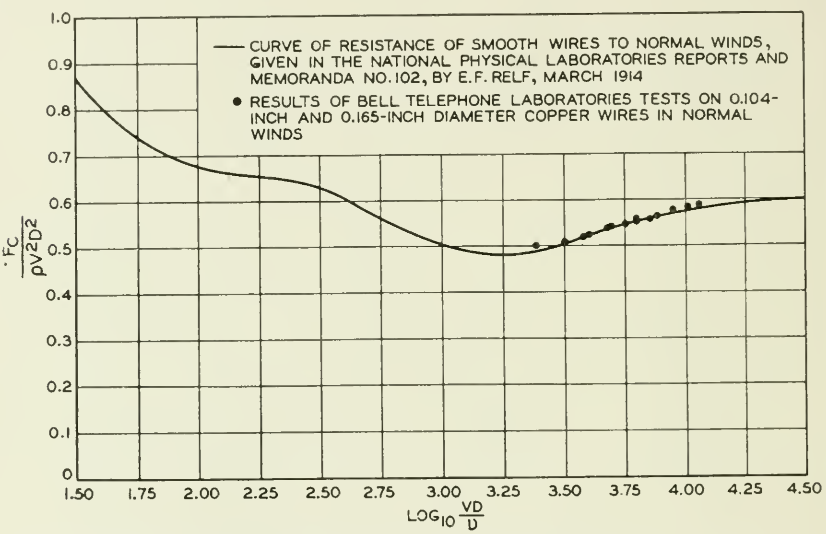


Fig. 5—Comparison of Bell Telephone Laboratories results for normal winds with the National Physical Laboratories (British) results.

the results with this in mind the following equation was obtained,

$$F_c = \rho V^2 \cos^2 \alpha D^2 f\left(\frac{VD}{\nu}\right).$$

This form of expression holds over the range of Reynolds number (VD/ν) covered and, also, over the range of angles from 0° to 60° from normal.

Since ρ (air density) and ν (kinematic viscosity of air) are constant for a particular atmosphere this equation becomes

$$F_n = K(V \cos \alpha)^2 D. \tag{6}$$

Here, F_n (the normal component of wind force) is measured in pounds per foot of wire. The diameter (D) of the wire is in inches and the actual wind velocity (V) is in miles per hour. This is the familiar equation for the force of normal winds with the addition of the term, $\cos^2 \alpha$.

Values of the constant (K) found for each value of the actual velocity (V) and angle are given in Table I (0.104-inch wire) and Table II (0.165-inch wire). The arithmetical averages ($\bar{X}$) of the constants for angles up to and including 60° and for angles up to and including 80° in the case of each velocity (V) are also given in these tables. As in the case of normal winds and as indicated by equation (6) K varies with the product of velocity and wire diameter (VD).

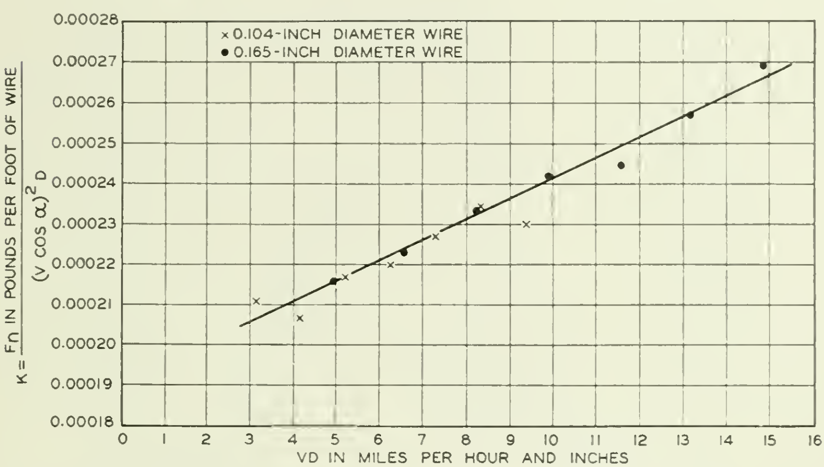


Fig. 6—Values of the Constant (K) in Equation $F_n = K(V \cos \alpha)^2 D$ for variations of the angle (α) of wind direction from normal—range, 0° to 60°, inclusive.

The relation between the average constant (K) for angles up to and including 60° and the product of velocity and wire diameter (VD) is summarized in the accompanying Fig. 6.

While our interest was centered mainly in evaluating the normal force of an oblique wind as stated above, some consideration has been given to the tangential force of an oblique wind and the variability of the angle between the normal and resultant wind forces.

The tangential forces of the oblique winds were determined by equation (3). Curves, for both sizes of wire, of tangential forces plotted against the angle α are given on Figs. 7 and 8 for all velocities (30 to 90 m.p.h.) used in the tests. The tangential force, of course, is a relatively small quantity as compared to the normal force. For this

reason some inconsistencies in the data and non-uniformity in the curves might be expected, particularly when plotted to such a large scale as used in these graphs. In Reports and Memoranda of the National Physical Laboratories [2] it appeared from their results that this force was not only small but fairly constant. The results of the tests reported here indicate that while the force is low in magnitude, it varies with the obliquity and the velocity of the wind and the diameter

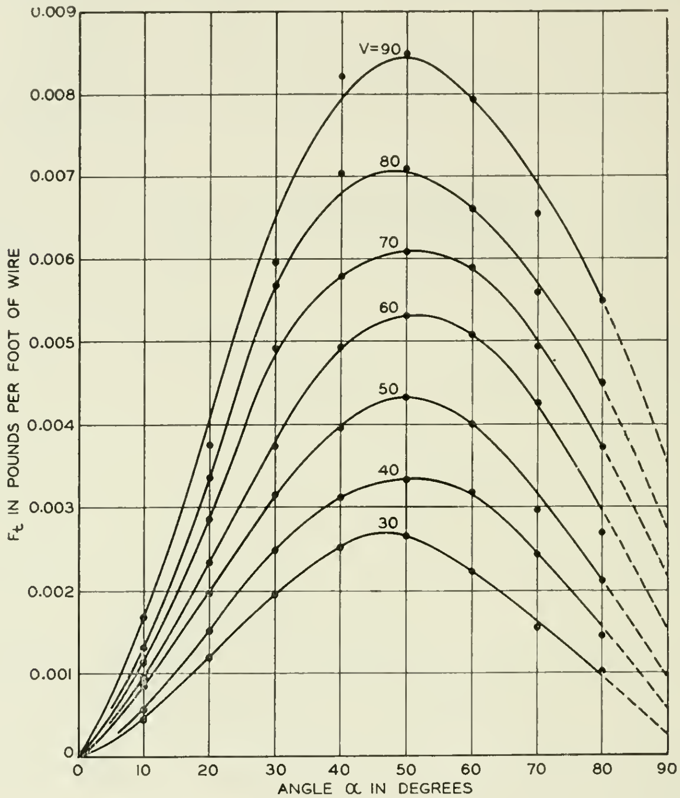


Fig. 7—Graphic relation between the wind force along the wire (tangential-F_t) and the angle (α) of wind direction from normal—0.104-inch diameter wire.

of the wire. In the case of 0.104-inch wire it increases from zero until the angle α (between the wind and the normal to the wire) is about 50° and then decreases as this angle increases. The action is similar in the case of 0.165-inch wire except the maximum is reached when α is about 60°. Whether this shift in the maximum with the wire diameter is real, and how far it would continue, is not clear since only two diameters of wire were tested. For 0.104-inch wire the variation in the force with

wind velocity when α equals 50° ranges from 0.00265 to 0.00850 pound per foot of wire for a range of velocities from 30 to 90 m.p.h. In the case of 0.165-inch wire and an angle (α) of 60° the variation ranges from 0.00254 to 0.00969 pound per foot of wire for the same range of velocities. As mentioned above the frame in which the wires were

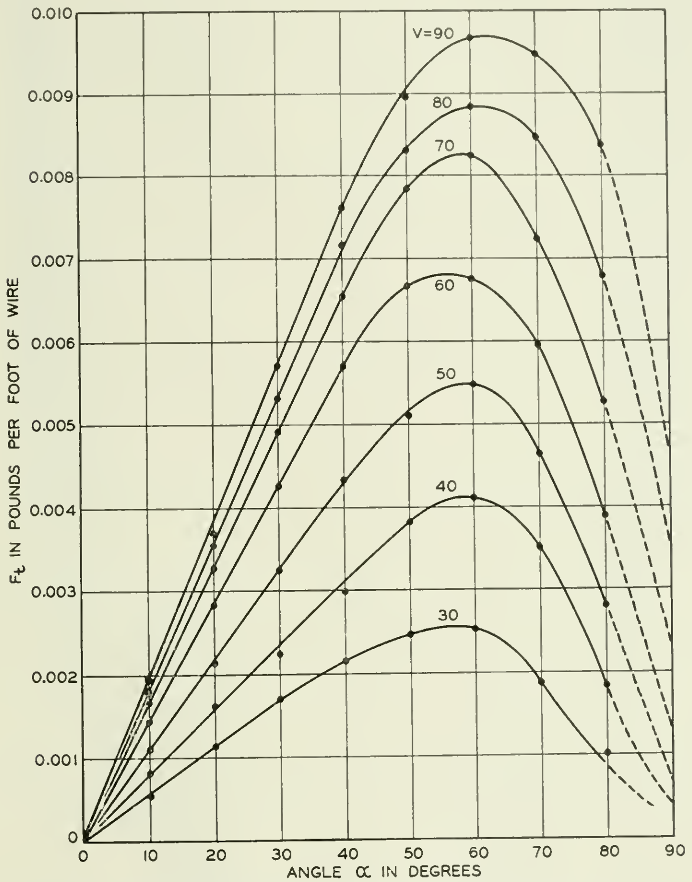


Fig. 8—Graphic relation between the wind force along the wire (tangential-F_t) and the angle (α) of wind direction from normal—0.165-inch diameter wire.

tested comprised such a major portion of the total resistance when the angle α approached 90° or the wind and wires were about parallel that the data for these cases were not considered reliable. In plotting the curves in Figs. 7 (0.104-inch wire) and 8 (0.165-inch wire) the tangential force for 90° was estimated and to indicate this the curves are dotted between the angles α of 80° and 90°.

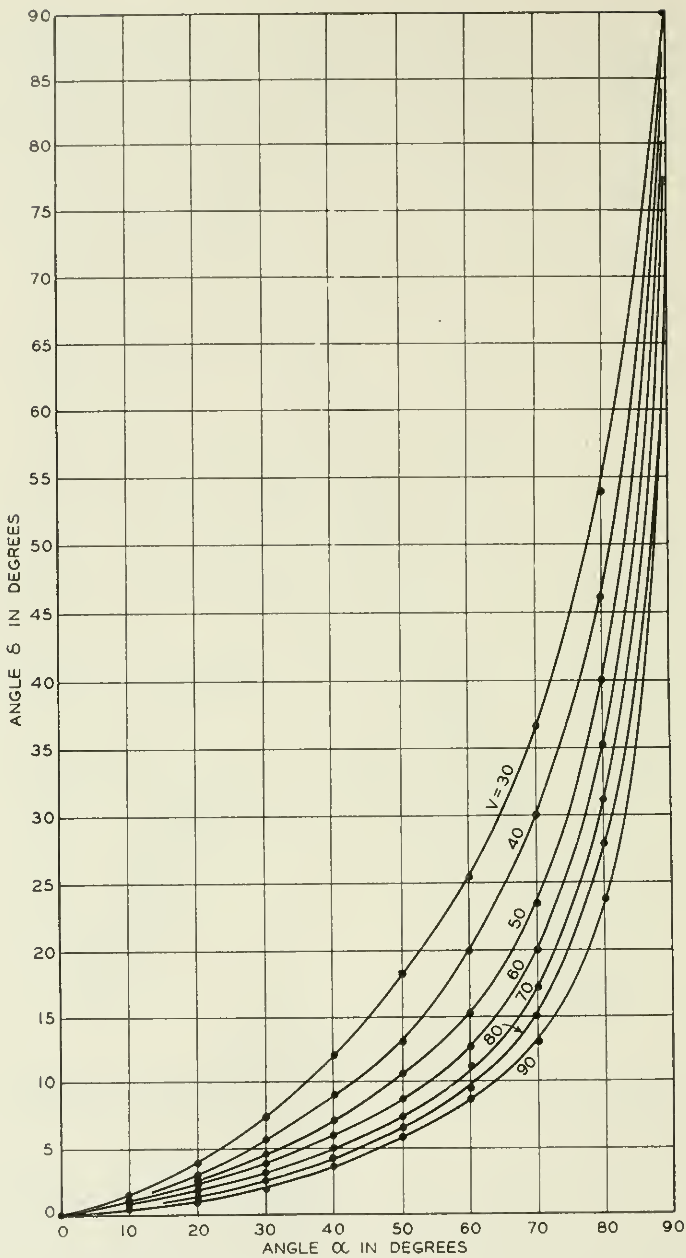


Fig. 9—Graphic relation between the angle (δ) of direction of resultant force from normal to the wire and the angle (α) of wind direction from normal—0.104-inch diameter wire.

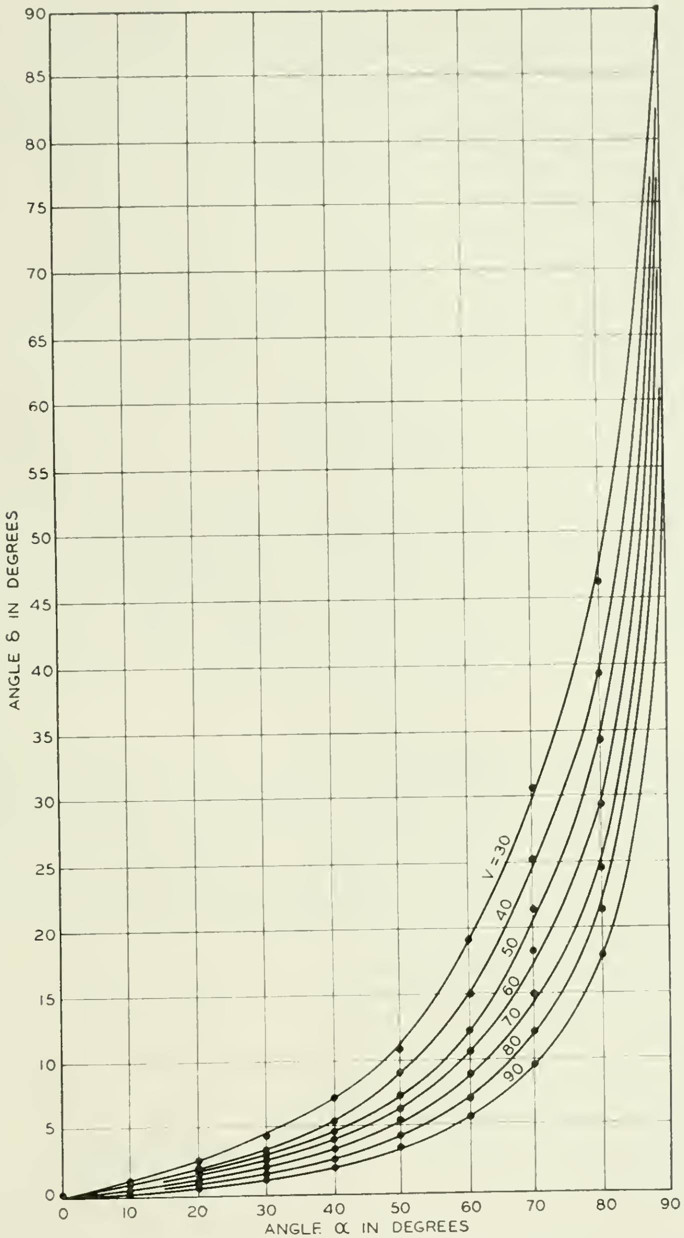


Fig. 10—Graphic relation between the angle (δ) of direction of the resultant force from normal to the wire and the angle (α) of wind direction from normal—0.165-inch diameter wire.

The angle (δ) between the resultant and normal wind forces was determined through the use of equation (4). The variations in this angle with the obliquity of the wind or the angle α are given for both sizes of wire in Figs. 9 (0.104-inch wire) and 10 (0.165-inch wire). From these graphs the following relations appear to exist in this range:

(*a*) For a given angle α the magnitude of the angle δ is inversely proportional to the product of velocity and wire diameter (VD),

$$\delta = \left(\frac{K_1}{VD} \right)_{\alpha=\text{constant}}.$$

This relation can be written

$$\delta = \left(\frac{K_2}{\left(\frac{VD}{\nu} \right)} \right)_{\alpha=\text{constant}},$$

where VD/ν is the familiar Reynolds number.

(*b*) For each size of wire and a given actual wind velocity the angle δ increases with the angle α. Hence, $\delta = f(\alpha, V, D)_{V,\, D=\text{constant}}$. Since the Reynolds number can also be considered constant

$$\delta = \varphi(\alpha, VD/\nu)_{V,\, D,\, \nu=\text{constant}}.$$

Conclusion

These tests indicate that the normal force on a wire due to an oblique wind is proportional to the square of the resolved component of the actual wind velocity for angles up to 60° from the normal to the wire. The expression for the normal force per unit length of wire is $F_n = K(V \cos \alpha)^2 D$, where V is the actual wind velocity and D is the wire diameter. The tangential component is relatively small as compared to the normal component.

Corrosion of Metals—II. Lead and Lead-Alloy Cable Sheathing

By R. M. BURNS

This paper discusses the corrosion of cable sheathing in the aerial and underground cable plants. Corrosion does not appear to be a primary factor affecting the life of aerial cables; failure of these cables occurs usually from intergranular embrittlement and is minimized by the use of alloy sheathing. It is shown that corrosion of cable sheathing in conduit occurs by means of the operation of small corrosion cells on the surface of the sheath or by the leakage of current from the sheath to ground. The driving force of these corrosion cells arises from some chemical inhomogeneity in either the metal or the surrounding environment. The course and the character of corrosion is determined chiefly by the influence of the constituents of the environment on the operation of these cells. These constituents may be classed as corroding or protective;—the corroding including oxygen, nitrates, alkalies and organic acids, while the protective are silicates, sulfates, carbonates, soil colloids and certain organic compounds. Cable sheathing buried directly in soils is seriously corroded by differential aeration-cell action resulting from physical contact of relatively large soil particles and metal. In general it is concluded that corrosion of cable sheathing is influenced more by the nature of the environment than by the chemical composition of the metallic material. The incidence of corrosion of cable sheathing is small owing to the maintenance of non-corrosive chemical and electrical environments in the cable plant.

THE intricate cable network of the telephone system offers numerous opportunities for the occurrence of corrosion. The property damage resulting from perforation of the sheathing by corrosion and the attending costly interruption of service have served to make the prevention of cable failure a matter of primary concern. The relatively low incidence of actual corrosion failures can be attributed largely to the vigilance of the electrolysis engineers and the plant forces.

Cable sheathing is one of the largest single uses of metallic lead. In 1929 it exceeded even that employed in the manufacture of storage batteries and constituted about 27 per cent of the entire consumption in this country. In the past fifteen years over two million tons of lead have gone into the communications and power cable plants. In the Bell System alone there are about 180,000 miles of lead alloy covered cables, about forty per cent of which are underground. About 95 per cent of the total mileage of telephone wires is in cable, the proportion of open wire construction decreasing each year.

The earliest telephone cables were of the type employed in telegraph practice, the individual wires being insulated with rubber or gutta

percha and the core covered with a rubber or textile sheathing. The first lead-covered telephone cables were made by David Brooks, Jr., and were installed in the year 1880. These consisted of cotton-covered wires drawn into a lead pipe—a moisture-proofing compound of rosin and paraffin being forced afterward into the pipe and allowed to solidify by cooling. The Western Electric Company began the manufacture of lead-covered telephone cables in 1881. These were of the so-called "Patterson" type and employed cotton-covered wires drawn into a lead pipe after which melted paraffin charged with carbon dioxide under high pressure, was forced into the pipe and allowed to cool, forming thereby a solid cake of paraffin between the core and the pipe. Cotton-wrapped wire alone took the place of this structure in 1884 and some four years later paper began to be substituted for cotton. Beginning in 1882 telephone cables were sheathed with an alloy of 97 per cent lead and 3 per cent tin, which continued to be the standard composition for cable sheathing in the Bell System until 1912. The general adoption of the present standard alloy of lead with 1 per cent antimony in that year has afforded substantial economies and a sheathing of high resistance to fatigue cracking. Recently, a new development, lead hardened with 0.03–0.04 per cent calcium, has shown in laboratory tests some promise as a cable sheathing material. In England ternary alloys of lead with cadmium and tin or with cadmium and antimony have been proposed. Unalloyed commercial lead is the covering generally used for power cables.

The lead which best lends itself to the manufacture of lead-antimony cable sheathing is a high-copper, low-bismuth chemical grade of lead of the following nominal composition:

Silver	0.002 to 0.02%
Copper	0.04 to 0.08%
Bismuth	0.005% (max.)
Arsenic, antimony and tin together	0.002 (max.)
Zinc	0.001 (max.)
Iron	0.0015 (max.)
Lead (by diff.)	99.90

There is no evidence that the copper content of this lead has any significant effect upon corrodibility when used in 1 per cent antimony sheathing, although it does appear to be a factor in certain other uses. Indeed, chemical composition appears to be of lesser importance than environmental influences in the corrosion of cable sheathing. The prevention of corrosion failures is mainly a matter of providing and maintaining non-corrosive chemical and electrical environments.

While the choice of lead as a cable sheathing material was dictated primarily by physical requirements, notably its adaptability to extru-

sion, corrosion resistance has been a large factor undoubtedly in its successful use. It has long been recognized that lead is one of the least corrodible of metals. Its dull unreactive character is synonymous with inertness. Widely used in ancient times for water pipes, roofings, caskets, linings for public baths, etc., many specimens have come down to us in nearly perfect states of preservation. The Romans, for example, employed lead water pipes in fifteen standard sizes usually ten feet in length [1] and some of these pipes are said to be in use today. In the form of roofings many examples exist which are five centuries old. It seems likely that corrosion has been less destructive than war to the original lead roofs of medieval cathedrals and buildings. Once a protective film has formed on lead the metal may be preserved indefinitely if not physically disturbed. In the air this film is usually an oxide while in the case of underground burial the film which forms on lead may be a silicate or in some cases merely a film of hydrogen shielded by the presence of soil colloids. In other instances sulfates and carbonates exert a retarding influence. Whether or not a protective film forms depends largely upon the character of the environment to which the metal is exposed. Under unfavorable conditions, such as exposure to acetic acid vapors, strong alkalies or contact with large soil particles, lead may be readily corroded. Purity of the metal plays a minor role in corrodibility in the atmosphere although it may affect its behavior in soil waters and other electrolytes.

The widely different conditions of exposure which prevail in the aerial and underground cable plants make it desirable to consider them separately. Corrosion caused by stray electrical currents, since it occurs mainly in the underground plant, will be discussed under that heading.

Corrosion of Aerial Cables

Corrosion is not a primary factor in the life of aerial cables. Failure of these cables is usually due to cracking and confined to sections which are subjected to repeated stresses or in some cases to prolonged mechanical vibration.[2] It is now recognized that the nature of the environment affects the endurance of metals to such stressing and vibration, and the term "corrosion-fatigue" has been applied to the embrittlement and cracking which result from the simultaneous application of tensile and compressive stresses and corrosive media.

The resistance of lead to corrosion-fatigue is lowered, for example, by exposure to the atmosphere.[3] Evidently the protective oxide coating which forms on lead in the air [4] is not only ineffective in preventing intercrystalline fracture under repeated stressing, but actually constitutes an accelerating factor. The specific volume of lead oxide is

greater than that of the metal from which it is derived [5] and it has been suggested that the presence of the oxide provides some sort of leverage which aids embrittlement.[6] It is possible that differential aeration cell action may be involved for it is conceivable that the surface oxide film in the region of the grain boundaries is the most susceptible to rupture, producing thereby areas which are anodic to the adjacent unfractured surfaces.

Intercrystalline corrosion of lead may be produced in the laboratory merely by immersion of the specimen in a solution of nitric acid and lead acetate.[7] The attack in this case, as also in the case of the simultaneous action of tensile stress and corrosion, occurs along the grain boundaries leaving individual grains of lead which retain the characteristics of the original metal.[8] While intergranular corrosion of this type can be produced in lead of high purity, the rate of attack in a given medium is usually a function of the purity of the metal.

Exclusion of the atmosphere or the use of coatings of certain oils or grease have been shown to retard the rate at which lead is embrittled by corrosion-fatigue.[9] More practical means of minimizing the intergranular failure of cable sheathing lies in modification of the composition of the sheathing.

Alloying with 3 per cent tin, or 1 per cent antimony materially increases the resistance of lead to intercrystalline embrittlement.[10] The antimony alloy has a considerably greater fatigue resistance than pure lead as measured in a certain type of laboratory fatigue test [11] but decreases in time, or when the alloy is cold worked, owing to an agglomeration of the dispersed antimony particles which occurs particularly in the region near the grain boundaries.[12] Lead alloyed with 0.04 per cent calcium and suitably age-hardened has been shown in laboratory tests to have a much higher resistance to fatigue failure than the 1 per cent antimony alloy.[13] Certain ternary alloys of lead containing cadmium are said to possess marked resistance to fatigue failure.[14] More recently lead containing 0.1 per cent tellurium has been shown to be about 3-fold more resistant than ordinary lead to mechanical vibration.[15] It should be emphasized that all of these comparisons of fatigue resistance were made in laboratory tests and are not based on field experience.

From the foregoing it will be seen that for the most part the aerial cable plant does not present a serious corrosion problem. The importance of the unavoidable environmental influences on sheath embrittlement is minimized by the use of lead alloy sheathing together with proper methods of cable suspension. Other types of lead corrosion are rare in aerial cables.

Corrosion of Lead Directly Buried in Soils

It is not the practice of the Bell System to bury lead-covered cables directly in the soil without the use of a protective coating. Recognition of the corrosion hazard involved in such construction is one of the considerations which led to the use of conduit for the housing of even the first cables which were placed underground. The more recent actual experience of certain small users with soil corrosion has served to confirm the soundness of this practice. The idea that cable sheathing might be buried safely in direct contact with soils was suggested by the fact that lead had been widely used as water pipes.[16] Many miles of telephone cables accordingly were laid directly in the earth, notably in Indiana, and frequently with unfortunate results.

In certain sections where it is considered economical to bury cables in the ground, a coating has been devised for the protection of the sheathing against corrosion. This consists in wrapping the lead-alloy sheathed cable with asphalt-impregnated paper followed by one or more layers of jute impregnated with a preservative compound, and in some cases steel tape armoring over which there is wrapped a final layer of jute. The structure is flooded with asphalt before and after each serving of paper and each layer of jute. The steel tape is employed where there exists any danger of induction from power lines; it may be omitted in locations where there is little likelihood of trouble from this source.

Before discussing the corrosion of cables in conduit, which is the principal concern of the present paper, it will be of interest to review the results of corrosion studies which have been made on lead and lead-alloy sheathing materials buried directly in soils. In addition to the presence of soluble salts, the underground environment in this case involves direct contact of the metal with relatively large soil particles and aggregates—a markedly heterogeneous condition. These points of contact of soil particles and metal become areas of reduced oxygen concentration as compared with surrounding regions of the metal surface which are more freely accessible to the soil atmosphere. The resulting oxygen concentration cells with a driving force of approximately 100 millivolts provide one of the most important means by which metals corrode in soils. The use of conduit affords an effective barrier against soil action of this character. Silt deposits which sometimes occur on cables in conduit do not give rise to differential aeration action probably because under such circumstances cathodic polarization of the corrosion cells is maintained.[17] This inhibitive function of soil colloids has been observed recently in connection with a study of

the corrosion-fatigue of drill pipe [18] and appears to be an important factor in the retardation of corrosion in certain soils.

The most extensive soil corrosion test is that which has been carried on under the auspices of the National Bureau of Standards.[19] In this test, specimens of both ferrous and non-ferrous metals were buried in 48 different soils in various parts of the country. Commercial lead and lead containing 1 per cent antimony were placed in these locations and specimens of these have been removed from time to time. These studies have shown that lead and the lead-antimony alloy are corroded by most soils, but at lower rates than are ferrous metals [20]—losses in weight averaging but 10 per cent and depth of pitting about 25 per cent of those shown by the iron and steel specimens. After approximately 10 years exposure in 18 soils the lead-antimony alloy was found in the majority of cases to be slightly but definitely more corroded than was commercial lead.

A smaller but more intensive soil corrosion test on lead and certain lead-alloys has been carried on by the Bell Telephone Laboratories in five typical soils in the general vicinities of Lafayette and Monon, Indiana. Three grades of lead * and alloys of these leads with antimony in amounts from 0.8 per cent to 2.5 per cent and with 3 per cent tin were chosen for this test. The specimens consisted of flat plates one square decimeter in area prepared from metal which had been extruded in the form of tape. Before burial these plates were degreased with carbon tetrachloride and scoured with fine sea sand. Five specimens of each material were buried at each location in a horizontal position at a depth of two feet. After a period of four years the specimens were removed from the soil, and after removal of the corrosion products, the losses of weight and the maximum depth of pitting determined.

The values for loss of weight are represented graphically in Fig. 1, in which all of the metallic materials are compared in each of the soils. The arithmetical averages are shown in all cases by means of broken lines. The maximum depth of pitting results showed a close correspondence to the losses of weight. For example, it was least in the Plainfield fine sand and the Fox silt loam and greatest in the Miami silt loam. The only specimens perforated by pitting were the lead-tin alloy and these only in the last mentioned soil.

* The grades of lead employed in this test and in the sulfation tests described later in this paper are designated as: Corroding or A.S.T.M. Grade I, 99.94 per cent lead; Chemical or A.S.T.M. Grade II, 99.90 per cent lead; and Common or A.S.T.M. Grade III, 99.85 lead. The principal impurity in chemical lead is copper, and in common lead, bismuth. The term "corroding" applied to the high purity product arose in connection with its use in the manufacture of white lead; it does not imply greater corrodibility.

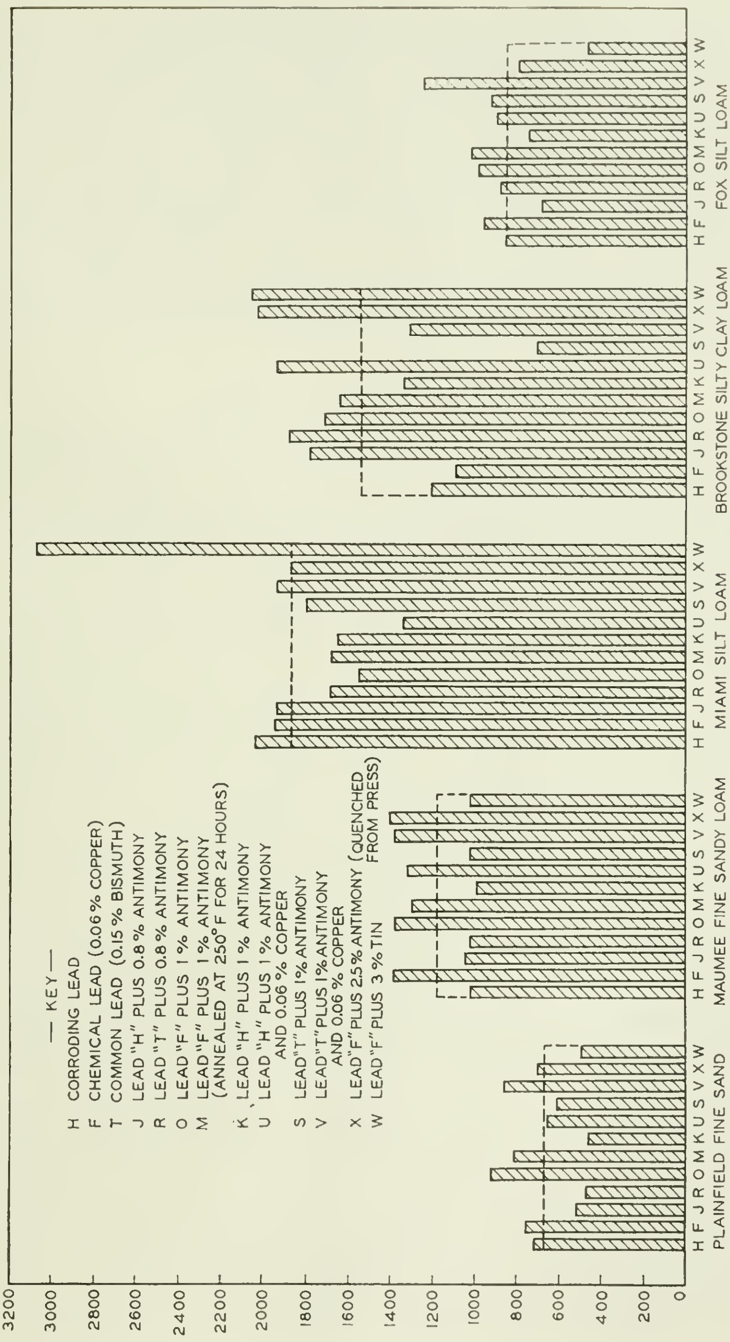


Fig. 1—Comparison of the corrosiveness of five soils toward lead and certain lead-alloys.

It will be seen that of the two variables, soil character and alloy composition, the former is decidedly the more important in its effect upon rate of corrosion. From an inspection of the data it would appear that there is no definite trend of corrodibility which may be correlated with composition. From a statistical analysis of the data obtained in this test it was concluded that variations in alloy composition within the scope of the test had no significant effect upon the corrosion behavior of the materials. In other words, the variations observed may be ascribed to chance and there is no indication of a significant difference in the rates of corrosion of lead and lead-antimony alloys when buried directly in the earth.

Corrosion of Cables in Conduit

The conduit mainly employed in the underground cable plant of the Bell System is a good grade of vitrified clay with glazed surfaces. Creosoted wood is widely used, particularly for single subsidiary cables. Wood has been employed extensively for main cables on the Pacific Coast where it offered an economical advantage. Steel or iron pipes find a limited application for certain special cases such as dips and relatively sharp bends. Heavy paper or fibre generally embedded in concrete has been used in a few instances. Concrete conduit has been employed by the utilities for power cables and in the telephone field to some extent abroad [21] but the danger of corrosion has militated against its adoption by the Bell System. It is possible that the greater heat dissipation of power cables as compared with telephone cables renders concrete conduit less hazardous for power cable use.

The environment to which underground cables in conduit are exposed is complex and varied. It is impossible to exclude moisture and soil air or vapors from the conduit. Surface waters may enter the cable compartments by way of the manholes and soil waters may seep through at duct joints or at small fissures which sometimes develop. The soil atmosphere tends toward higher concentrations of carbon dioxide and lower oxygen than the outside air; it is often high in humidity resulting in the condensation of drops of moisture on the sheathing. Acetic acid vapors arising from wood conduit or other sources may contaminate the duct air. Muddy soil waters may deposit a layer of silt on the sheathing. These waters contain varying amounts of salts, acids or alkalies. Free lime leached from concrete structures or caustic alkali produced, as indicated in the following paragraph, by the electrolysis of sodium chloride are the principal alkaline constituents. Even leakage from sewers is sometimes a contaminating influence which induces corrosion.

In addition to chemical influences, the electrical condition of the cable with respect to earth or to other adjacent metallic structures has an important bearing on corrosion. Where it can be done without jeopardy to other structures, it is desirable to maintain the cable network very slightly (of the order of 0.2 volt) negative or cathodic to earth. There is evidence that under this condition the sheathing is less readily corroded by couple action of a miscellaneous character. At appreciably higher negative potentials alkali or lime salts may be electrolyzed producing thereby caustic alkali or free lime which are corrosive to sheathing. On the other hand, an electrically positive condition of the cable may be conducive to the ordinary "stray-current" or anodic corrosion.

The Origin and Nature of Corrosion Cells on Cable Sheathing

The mechanism by which cable sheathing corrodes in conduit involves the replacement by the metal of hydrogen or another metal in compounds present in the surrounding environment—a process which has been described in some detail in a previous paper.[22] Most commonly, the dissolving lead replaces hydrogen from water. The areas or points on the sheathing at which lead dissolves are the anodes of small corrosion cells, the cathodes of which are the regions at which hydrogen is deposited. The driving force of these cells arises either from some chemical or physical inhomogeneity of the metal, or some inhomogeneity of the environment. Their electrolytic operation is influenced by the conductance and chemical nature of the environment, and by the size and distribution of the anodic and cathodic areas.

Corrosion cells owing their origin to sheath composition are exemplified by the presence of two metallic phases, one of which is lead and the other either an impurity, such as copper, bismuth, nickel, zinc, etc., or a hardening agent such as tin, calcium, cadmium, or antimony. Copper and antimony, for example, are cathodic to lead under the prevailing conditions and facilitate the discharge of hydrogen. The small proportion of cathodic area on the metal surface in both cases, however, will result in high cathodic current densities inducing polarization, and a low rate of corrosion except perhaps in acid solutions where the potential of the lead-hydrogen cell will be increased. This acceleration in acid solution is borne out by laboratory corrosion tests which show that the rate of corrosion of lead containing 3 per cent tin is about 50 per cent greater and lead containing 1 per cent antimony is about 10 per cent greater than that of soft lead in dilute (0.001 molar) acetic acid. Antimonial lead is said to corrode more rapidly than pure lead in humic acids.[23]

Relatively larger areas of copper deposited upon the cable by replacement from copper salts have been observed to promote corrosion of the sheathing. Scraps of copper wire corroded in manholes by saline waters are believed to have been the source of the copper compounds in such cases.

Wiping solder in contact with sheathing at the splicing sleeve provides still another example of a corrosion cell originating from the contact of diverse metals. Laboratory measurements of the potential of this couple in dilute chloride, alkali and acid solutions show solder is usually the anodic or corroding electrode. The observed potential differences in hundredth molar solutions at room temperature were as follows:

Solution	Potential Difference in Millivolts
Potassium chloride	6 ± 3
Caustic soda	11 ± 3
Acetic acid	20 ± 8

Similarly, the 3 per cent tin-lead sheathing in contact with 1 per cent antimony sheath would give rise to a galvanic couple in which the former would be anodic but by smaller values of potential than given above for the solder-sheath couple. Ordinarily in the soil water environments which prevail in the underground plant the potentials of neither of these couples is sufficient to maintain current flow and there is no evidence of attack. The few cases of corrosion of this type which have been observed, and which have been characterized by pitting of the solder and even of the sleeving (where this was 3 per cent tin) are believed to have arisen in electrolytes somewhat alkaline in nature which contained abnormally low concentrations of film-forming constituents such as silicates, sulfates or organic colloids.

In general, the influence of metallic composition upon corrodibility may be readily detected by measuring the rate of sulfation of the metallic material in sulfuric acid.[24] This test provides a method of measuring surface activity and affords a means of comparing the relative rates at which similar alloys tend to corrode in corrosive environments or tend to become passive in the presence of film-forming constituents. The sulfation-times measured by means of a recording potentiometer have been determined on specimens of leads of various compositions and for several cable sheath alloys in 7-normal sulfuric acid. The averages of four determinations made on each material bore the following relationship to each other, assuming the sulfation time of spectroscopically pure lead to be one hundred:

Spectroscopically pure lead (99.999% lead)	100
Corroding lead (A.S.T.M.—Grade I, 99.94% lead)	80
Chemical lead (A.S.T.M.—Grade II, 99.90% lead, contains 0.06% copper)	65
Common lead (A.S.T.M.—Grade III, 99.85% lead, contains 0.13% bismuth)	70
Corroding lead alloyed with 1.5% tin, and 0.25% cadmium	85
Chemical lead alloyed with 1.5% tin and 0.25% cadmium	75
Corroding lead alloyed with 3% tin	70
Chemical lead alloyed with 0.04% calcium	80
Corroding lead alloyed with 0.04% calcium	75
Common lead alloyed with 0.04% calcium	55
Corroding lead alloyed with 0.5% antimony and 0.25% cadmium	25
Chemical lead alloyed with 0.5% antimony and 0.25% cadmium	25
Chemical lead alloyed with 1.0% antimony	20

From an inspection of these results it appears that the surface reactivity of lead is markedly increased by the presence of impurities or by alloying with small amounts of other metals. Of the hardening agents chosen for study, tin, tin and cadmium, and calcium exert the smallest influence on rate of sulfation, while antimony, whether used alone or with cadmium, has the most pronounced effect. The presence of small amounts of copper appears to have an accelerating effect upon reactivity as does also the presence of bismuth. This adverse effect of bismuth has been noted in connection with the use of lead in sulfuric acid plants.[25] Since the environment in which cables are used contains both corrosive and film-forming substances this comparison of rates of sulfation of lead and its alloys is not necessarily a direct indication of the relative rates of corrosion of these materials when used as cable sheathing.

It is well known that the intensity with which metals tend to ionize is affected by their physical state, small crystals and strained structures possessing higher intensities and therefore more electronegative or anodic potentials than large crystals and annealed structures. In the case, however, of lead and most lead-alloys suitable for cable sheathing, self-annealing occurs at ordinary atmospheric temperatures, and for this reason it is highly improbable that corrosion is ever initiated as a result of physical condition of the metal.[26] In the laboratory it was found that lead intensively worked at liquid air temperatures, where self-annealing does not occur, was from 2 to 3 millivolts electronegative to annealed lead when measured immediately afterward at 25° C. in 0.2 normal lead-acetate solution. This potential difference was reproducible but could not be maintained for more than ninety minutes at room temperature.

It is conceivable that scratching or mechanical injury of cable sheathing, such as might occur during installations, could give rise to the familiar metal-metal oxide corrosion cell. The operation of this cell has been demonstrated in a laboratory experiment in which pieces

of lead covered with litharge were freshly scratched after being submerged in water.[27] It is of significance that although corrosion readily occurred in this experiment, there was no attack if the scratch were exposed to the atmosphere for two hours before submersion of the specimen. In other words, the oxide film on lead is readily self-healing and injury to it is unlikely to cause corrosion.

Turning to a consideration of corrosion cells originating from the exposure of the sheathing to an inhomogeneous environment, reference has already been made in discussing soil corrosion to the nature and the importance of oxygen concentration cells, and to the protection which the conduit affords against this hazard. Cables in conduit are seldom subject to contact with the character of inert objects which lead to the establishment of oxygen concentration cells. Relatively large hard particles are generally the most effective agents in producing differential aeration. In the laboratory, lead can be pitted by contact with a glass rod when submerged in a dilute sodium chloride solution. There are a few instances in which cables in conduit appear to have corroded by means of oxygen concentration cells. For example, there is evidence that deep pits in sheathing produced by the leakage of stray currents to earth have continued to deepen to the point of perforation of the sheathing, after removal of positive potential conditions. The bottoms of such pits are less accessible to oxygen and appear in some cases to function as the anodic elements in differential aeration cells. Cases of this kind are generally diagnosed by the field forces as "old action."

Another, but rather uncommon, example of oxygen concentration cell has been observed in the use of a porous duct plugging material contaminated with acetic acid. In this case it seems likely that the naturally protective oxide film on the sheathing was destroyed by the acid following which this region, owing to the exclusion or partial exclusion of oxygen, became anodic to the adjacent areas which were freely accessible to air. Contamination with acetic acid does not appear to be essential to this action since other cases have been reported in which the duct plugging material was free from acid.

Finally there exists the possibility of large scale differential aeration cells where one cable of a multiple run is placed, owing to space limitations, in a dip under a large sewer, but bonded to the other cables. Such a cable may suffer severe corrosion in the region of the dip as a result of the lower oxygen content of the atmosphere in this duct as compared to that prevailing in the other ducts.

The discussion of differential environments has related so far only to oxygen concentration. In a similar manner, underground cables

may be exposed to different hydrogen ion, metal ion or salt concentrations and where the demarcation between concentration zones is sufficiently pronounced may give rise to differential concentration cells, the driving forces of which have theoretical values of 29.5 millivolts per ten-fold difference in ion concentration. An examination of the Rhineland cable which connects Berlin and Cologne has shown that the most extensive corrosion occurred at points where there was an abrupt change in the character of the soil or geological structure.[28] There is usually sufficient diffusion and circulation of underground waters to equalize ionic concentrations and prevent the development of cells of this type assuming serious proportions.

Effect of Environment on Operation of Corrosion Cells

The nature of the more common electrolytic cells by means of which cable sheathing corrodes has been discussed at some length. Consideration will now be given to the manner in which various environmental factors influence the operation of these cells. In general, these factors may be classified either as corroding or film-forming agents, although their influence will depend quite as much upon their concentration as upon their specific nature. It is meaningless to report that a metal corrodes or does not corrode in this or that electrolyte unless full experimental details are given. Only with a complete knowledge of the condition of a metal surface and of the nature and concentrations of the components of its environment can the resulting behavior be predicted. For example, it has been shown that it is often the ratio of the concentrations of film-forming to corroding substances which determines the character of attack.[29] For high values of this ratio, the metal will be protected; for low values it will be uniformly corroded, but for intermediate values of this ratio, the surface will be only partly protected with the result that corrosion will be localized in the form of destructive pitting. With these limitations in mind, some of the principal constituents of the environment which affect the behavior of cable sheathing may be classified as follows:

Corroding	*Protective*
Oxygen	Silicates
Nitrates	Sulfates
Chlorides	Carbonates
Alkalies	Colloidal substances
Organic acids	Certain organic compounds

Of the corroding elements, oxygen is the most important in its effect upon the operation of corrosion cells. Owing to the high potential required to discharge hydrogen on pure lead (i.e., its high hydrogen over-voltage) these cells tend to cease functioning owing to

cathode polarization. The role of metallic impurities of lower overvoltage in discharging hydrogen has already been considered. Oxygen, it is obvious, aids corrosion by depolarizing cathodic areas on the surface of the sheath. That the effect of oxygen is proportional to its partial pressure in the atmosphere has been found in a laboratory study, the results of which are given in Fig. 2. In this experiment six specimens of extruded chemical lead, each of an area of one square decimeter, were prepared for test by degreasing with carbon tetra-

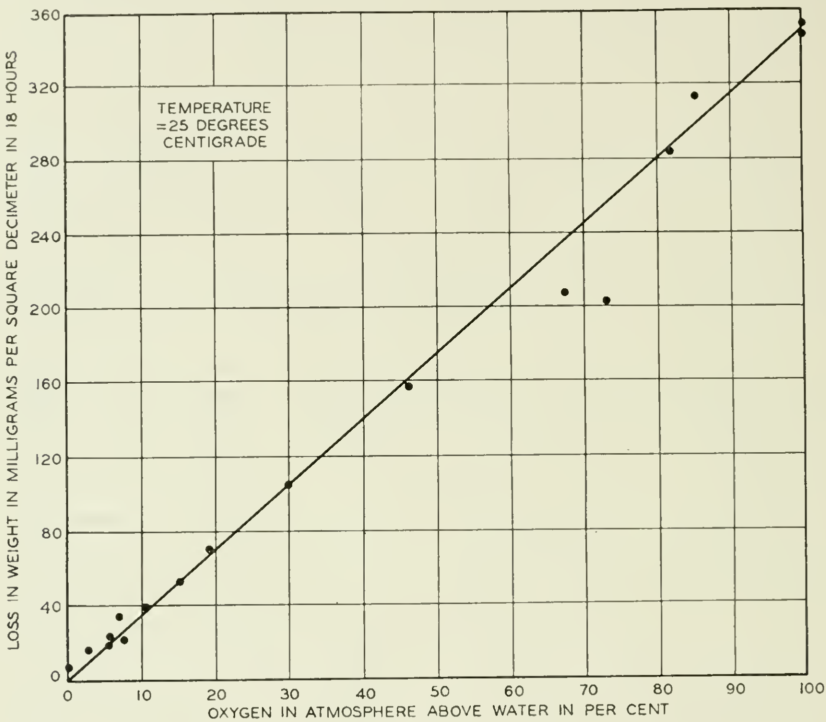


Fig. 2—Effect of oxygen on corrosion of lead submerged in distilled water.

chloride, and the tarnish film removed by dipping in dilute acetic acid (one part of acid to five parts of water). These specimens washed and dried, and weighed to the nearest milligram, were submerged in large jars of distilled water which had been previously saturated with purified nitrogen, oxygen or various mixtures of the two. The corresponding atmospheres were maintained above the surface to the water during the test. After a period of 18 hours the specimens were removed, washed and the losses of weight determined.

A number of cases of cable sheath corrosion have been attributed to the presence of nitrates in the duct electrolyte. The nitrate content of most soil waters is very low—a few parts per million ordinarily—but occasionally contamination from industrial plants or from sewers, the organic matter from which may undergo nitrification, has led to a several fold increase. Concentrations of nitrate from 20 to 425 parts per million have been found in the electrolyte from locations where failure of the sheathing occurred. It has been shown that solutions containing a thousand parts of nitrates per million are markedly corrosive to lead.[30] Another investigator reports that the addition to natural soft waters of nitrates in excess of 50 parts per million increases their corrosiveness about 20 per cent, higher quantities of nitrate being required to produce this effect in hard waters.[31] Often the corroded region of the sheath is black in appearance owing to the presence of loosely adherent, finely divided lead (possibly oxidized) and antimony which accumulates by some sort of undermining action during the rapid attack. The formation of this black coating in the presence of nitrates is reported by others.[32]

The mechanism of nitrate action at the cathodes of corrosion cells is similar to that of oxygen and of lower concentrations of oxidants in general, and consists in depolarization. In addition, the high solubility of lead nitrate prevents an appreciable polarization of anodic areas. It is of interest to note that in the presence of nitrates, in the form of nitric acid, oxygen furnishes but little additional acceleration of corrosion. For example, in 30 per cent nitric acid, the ratio of the rate of corrosion in the presence of oxygen to the rate in the absence of oxygen has been shown to be 1.1, while in 20 per cent hydrochloric acid and glacial acetic acid, this ratio is 10.0 and 10.9, respectively.[33]

Cable sheathing is little affected by the chloride content of most ground waters. In tests, for example, in which chlorides were added to natural waters, there was no increase in corrosion when the chloride content was of less than 1000 parts per million, a value seldom attained in soil waters.[34] Even infiltration of sea water does not constitute a corrosion hazard unless the cable is markedly negative to earth. Indeed, in sea water the corrosion of lead may be retarded by an encrustation of lead chloride which forms on the surface of the metal, as well as by the lower prevailing oxygen content. Extruded bars of lead and lead containing 1.6 per cent antimony, 60 cm. in length and 2.87 cm. in diameter exposed for four years to tidal action have shown losses in weight of 0.65 per cent and 0.51 per cent, respectively,[35] but doubtless mechanical erosion was an important factor in this rather drastic test. Laboratory tests made on lead foil in a sodium chloride solution show

that the rate of corrosion increases with increasing salt concentration up to a maximum at 1 per cent and that at concentrations of 3 per cent, which corresponds roughly to that of sea water, the rate is markedly less.[36] The favorable experience with cables submerged in sea water or mixtures of sea water and soil waters indicates that corrosion inhibitive agents exert under these exposures a predominating influence. The effect of the chloride content of the duct electrolyte must be mainly one of increasing the conductivity, there being insufficient concentrations in relation to the concentration of film-forming substances to produce even pitting or local attack.

A type of cable sheath corrosion of considerable importance is that which is fostered by alkalies. It is characterized usually by the formation in the region of attack of deep red crystals of lead monoxide or litharge. Occasionally the yellow form of litharge or a greenish hydrated lead monoxide may appear, but in one case where the strength of caustic was so great as to cause discomfort upon handling the cable, no colored compounds developed. The red monoxide crystallizes out of saturated solutions of alkali plumbites which are formed by the solution of lead in alkalies.[37] It can be produced in the laboratory by immersing specimens of lead in saturated lime water and aerating the solution for several days with carbon dioxide-free air. The appearance of this red oxide on cable sheathing is a certain indicator of alkali attack. If detected before failure of the cable, the action can be stopped usually by removing the source of the alkali and thoroughly flushing the cable conduit with water.

A source of alkali affecting underground cables is concrete conduit and occasionally other concrete structures. Free lime in the surface layers of fresh concrete is usually converted by the action of carbon dioxide into calcium carbonate within a few weeks and this is less alkaline in nature. Seepage of moisture through concrete which may occur in less dense grades of this product may leach free lime from within. At the Panama Canal water seeping through the concrete floors and walls of lock chambers caused serious corrosion of the sheathing of power cables in a short time. Analysis of the seepage water disclosed high alkalinity. In the same locality a telephone cable in vitrified clay conduit was corroded by the seepage of water through cement sacks used for wrapping the conduit joints.[38] Greater attention is being given recently to the production of concrete conduit of greater impermeability, lower alkalinity and to "curing" methods. The use of high alumina cement, which is much less corrosive to cable sheathing,[39] has been proposed.

Another important source of alkali is the electrolysis of sodium chloride or common salt by electrical currents flowing to the cables. Under these conditions caustic soda is produced at the sheath which is cathodic or negative to earth; hence the terms "cathodic" or "negative" corrosion. The salt usually comes from that used in the winter to thaw out street car switches, although in one case it has been traced to the drippage from salt-ice mixtures of ice-cream delivery trucks. In still another case, a power cable, negative to earth, suffered alkaline attack as the result of the electrolysis of alkali salts concentrated at a low point in the cable run by heat dissipation of the cable. Finally it should be mentioned that the use of calcium chloride on streets for melting snow or laying dust would lead undoubtedly to its coming in contact with the underground cable plant and being converted into corrosive free lime in areas negative to earth.

It has been known since the Middle Ages that lead is corroded by acetic acid. In the presence of the carbon dioxide of the atmosphere, the corrosion product is the pigment, white lead. The attack manifests itself by the formation of a white encasement around the globules of moisture on the sheath; at first a mottled effect is produced which in time develops into a heavy white encrustation of the carbonate or basic carbonate of lead. The early use of wood conduit was attended with occasional cases of acetic acid corrosion and it was found that the wood tar creosote used as the preservative contained this acid. Since that time coal tar creosotes have been specified for the preservation of wood conduit. The conduit most widely used in this country is yellow pine. Properly creosoted this product has not been known in Bell System experience to cause corrosion except when used under such unusual circumstances as close proximity with steam pipes or exposed on viaducts over railroad yards to the heat of locomotive stacks. In these cases acetic acid was liberated as a product of the slow decomposition of the wood. A recent instance of acetic acid attack in creosoted conduit manufactured from southern yellow pine has been reported and attributed to acid liberated by the destructive decomposition of the wood by the Kansas sun.[40]

The most serious corrosion of cables by acetic acid on record is that which occurred on the Pacific Coast in creosoted Douglas fir conduit a few years ago.[41] Following the initial satisfactory use of this product for subsidiary cables it was employed extensively for main communication subways. With the expansion of the cable plant into this newly constructed duct system several cases of acetic acid corrosion occurred—most of them within the first 15 months in conduit of recent installation. Analysis of the atmosphere within the cable compart-

ments revealed the presence of corrosive concentrations of acetic acid. In the investigation made of this trouble it was concluded that the high native acidity of Douglas fir, together with the drastic treatment required to impregnate it with creosote, offered a reasonable explanation for the corrosiveness of the conduit.[42] The corrosive action was effectively stopped by neutralizing the acid with ammonia gas supplied to the affected conduit in a 2 per cent mixture with air.

The corrosiveness of air laden with acetic acid vapors lies in the persistence of effective non-polarized corrosion cells of constant voltage. The acid furnishes an abundant and reasonably constant source of replaceable hydrogen ions and the continued precipitation of lead as carbonate by the action of carbon dioxide maintains a low concentration of lead ions. Oxygen acts as a cathodic depolarizer. Since the precipitation of lead carbonate or basic carbonate occurs at an appreciable, although very small, distance from the seat of activity on the metal surface, it offers little or no hindrance to the corrosion action.

Phenols and other acidic constituents of coal tar pitches have been reported to be corrosive to cable sheathing when in direct contact in the form of protective coatings.[43] There is no evidence either from experience with creosoted conduit or from laboratory tests that phenolic vapors from creosote are appreciably corrosive to sheath.

So much for the corrosive media of the environment of the underground cable plant. Of the protective agents, none is more important than soluble silicates. It is well known that lead is markedly corroded in distilled water, and by waters low in hardness and in total solids. Saturation of distilled water with calcium silicate (soluble to the extent of about 100 parts per million), or with silicic acid derived from a suspension of silica flour, will prevent corrosion of lead. The corrosiveness of certain natural waters has been greatly reduced by the addition of only 10 parts of sodium silicate (expressed as silicic acid) per million.[44] Analysis of a large number of samples of waters from cable manholes and subways has shown silicate contents of from 2 to 25 parts per million. In concrete conduit values up to 143 parts per million have been found. It is of interest in this connection to note that although silicates appear to protect lead to some extent in all ground waters, their effectiveness is greatest in the range of alkalinity corresponding to values of pH between 9 and 11, where pH equals the logarithm of the reciprocal of hydrogen-ion concentration. The resistance of underground cables to corrosion appears to depend chiefly upon the film-forming action of silicates. The minimum concentrations required to give protection depend upon the nature and concentrations of the corroding agents which are also present.

The effectiveness of silicates in passivating lead lies in the extremely low solubility of lead silicate. Consequently silicate ions are precipitated as lead silicate in close contact with the sheath at the anodic areas of the corrosion cells. As the more anodic regions become polarized in this fashion other areas tend to function as anodes but with the same result until the surface of the sheath becomes entirely covered with an insoluble coating of lead silicate which is impervious to the corrosive elements of the environment.

Chromates and phosphates stand next to silicates in ability to passivate lead, but do not occur in the electrolytes in contact with underground cables. Sulfates, however, are a common constituent of these environments and in laboratory studies have been shown to be as effective as phosphates.[45] The passivating effect of sulfates is directly proportional to concentration, 2500 parts per million reducing the rate of corrosion of distilled water about 50 per cent.[46] Electrolytes from the cable plant seldom contain as much as 10 per cent of this amount of sulfate and so the specific contribution of sulfates alone is not large; however, added to that of various other film-formers it is of importance.

Carbonates exert a marked retarding influence on the corrosion of lead. The water which comes in contact with underground cables always contains carbonate ions derived either from soluble carbonates from the soil or from carbon dioxide of the soil atmosphere. Numerous analyses of the air in cable ducts has shown it to run from 0.1 per cent to 10 per cent of carbon dioxide, usually averaging about 1.5 per cent or 0.015 atmospheres pressure. Pressures of carbon dioxide within this range reduce the rate of corrosion of lead in distilled water about 50 per cent. It is claimed that high pressures of carbon dioxide, e.g., 6 atmospheres, increases the solvent action of water on lead.[47] Carbonate equilibria calculations of the system lead carbonate-carbon dioxide-water show that the film of corrosion products which forms on lead in aerated distilled water is a hydrated oxide of lead when the partial pressure of carbon dioxide is less than 10^{-14} atmospheres. Above this value for carbon dioxide and up to a pressure of about 10 atmospheres, the film should consist of lead carbonate. Basic carbonate, if a true solid phase, should also be found within this range. The bicarbonate of lead would appear to be stable at still higher carbon dioxide pressures. It is of interest that there is a minimum in the calculated solubility curve for lead carbonate in the region of 10^{-6} atmospheres of carbon dioxide. Increasing solubility at pressures greater than this is due to the increasing concentration of bicarbonate ions. This means that the effectiveness of soil carbonates in passivating cable sheathing is somewhat reduced by the higher carbon dioxide

pressures which obtain underground. It is still, however, one of the most important of corrosion inhibitors.

Mention has already been made of the protective influence under certain circumstances of silt or clay deposits on the surface of the sheath. There are in the underground electrolyte many other substances mainly organic in nature and often colloidal which aid in the preservation of cable sheathing. Whereas the anions, such as silicates, sulfates, and carbonates, induce passivity by a process of anodic polarization of corrosion cells, the inhibitive mechanism of soil colloids and of the organic materials in soil electrolytes is usually one involving cathodic polarization of these cells.

Stray Current Corrosion

The most common kind of cable sheath corrosion, the most destructive and best recognized, is that which occurs when electrical currents flow from the sheath to ground. In this case the portion of cable of higher potential than earth has the general characteristics of an anode, while the cathode is some extraneous structure. The potentials between anode and cathode may be and generally are greater than those which are possible for the electrolytic corrosion cells which have been described at length in this paper. The size of the currents which may flow for a given potential will of course depend upon the resistance of the path, i.e., upon the electrolytic resistance of the soil solution in contact with the cable. The size of the anodic area will depend upon the area of the sheath in contact with the electrolyte. The nature of the corrosive attack accordingly will depend upon this area and the rate of current flow or, in other words, the current density. In appearance the corroded area may be a clean cut pit or pits, or it may be roughly etched. When the potential is greater than about 2 volts, a brown colored anodic oxidation product, lead peroxide, may be formed. A simple test for this—the blue coloration which develops when a small amount of it is dissolved in a 5 per cent solution of tetramethyldiaminodiphenylmethane containing dilute acetic acid—is a certain indicator of anodic action. A negative result with this test, which is the more common experience, does not, however, exclude the possibility that the attack was anodic in character; the potential to earth may have been too small or the peroxide may have been actually formed but may have been consumed by local action following removal of the positive sheath potential.

Occasionally lead chloride, a white salt, may be formed in the corroded areas under anodic conditions. Thus, the finding of a relatively greater concentration of chloride in the corrosion product

than in the surrounding environment is trustworthy evidence of anodic corrosion.

The corrosion efficiency of stray current anodic corrosion, i.e., the per cent of the current involved in dissolving lead, will often be appreciably less than 100 per cent. The complementary anodic reaction occurring at voltages greater than approximately 2 volts is the evolution of oxygen. For example, in an extract of a black alkali soil containing high concentrations of sulfates, tests showed that less than 1 per cent of the current was consumed in the dissolution of lead. Under the conditions generally prevailing, however, it is likely that the corrosion efficiency is reasonably high and that the amount of corrosion will be nearly proportional to the amount of current which flows from the sheath to ground.

Cathodic or negative corrosion of cable sheathing, which occurs when current flows from earth to the sheath, has been described already under the discussion of alkaline corrosion. A not uncommon indication of negative conditions is an encrustation of calcium carbonate on the cable. In this case the sheathing is generally not corroded. It seems likely that the alkali produced by electrolysis of lime salts is carbonated as formed and before reaching sufficiently high concentration to initiate corrosion and that calcium carbonate so formed crystallizes on the surface of the sheathing.

Instances have been observed in which cable sheathing appeared to have corroded from the inside surface.[48] It is believed that the action in these cases was preceded by the occurrence of cracks or fissures in the sheath which admitted moisture and provided electrolytic paths by means of which current flowed from the sheathing to the copper conductors within.

Destruction of cable sheathing by stray electrical currents derived from large-scale galvanic cells has been experienced. In this case, which at first was rather mysterious, it was found that contact of iron pipes with beds of buried cinders set up large iron-carbon couples with potentials of approximately 0.7 volt. The soil at this location was unusually low in resistance and the wood conduit in which the cables were housed was water-logged with the result that the cables picked up current in regions near the iron structures and lost it at other points where the cable passed through the general neighborhood of the cinder beds. The electrical condition of the cables, determined by pulling through an adjacent duct a modified calomel reference electrode,[49] showed that the potential of the cable with respect to earth varied sharply from point to point and often reversed itself more than once in a section between two manholes. Removal of corroded cables

always confirmed the duct survey as to the location of anodic regions. The trouble was corrected by removal of the cinder beds.

Summary

In summarization it may be stated that corrosion is not a primary factor affecting the life of aerial cables. The tendency of lead to crack as a result of repeated stressing has been minimized by alloying with one per cent of antimony, and aerial cables sheathed with this alloy have shown satisfactory resistance to embrittlement of this character. When cable sheathing materials are buried in direct contact with soils, serious corrosion develops as a result of differential aeration cell action, and appears to have little or no relation to chemical composition of the metallic material. In addition to corrosion cells which originate in some inhomogeneity of the environment, such as the partial exclusion of air, corrosion of cable sheathing may occur by means of galvanic cells arising from the presence of metallic impurities or contact with a more noble metal such as copper. The electrolytic operation of these corrosion cells is influenced by the conductance of the surrounding electrolyte and the chemical nature of its components. Such constituents as oxygen, nitrates, alkalies, organic acids and chlorides (in low concentrations) facilitate the operation of these cells, thereby increasing the rate of corrosion, whereas silicates, sulfates, carbonates, colloids and certain organic compounds of the soil waters exert a protective action which may retard or prevent corrosion. Finally mention is made of the characteristics of the most common kind of corrosion, that due to stray electrical currents. This may occur as anodic action when current flows from the cable to earth, or it may occur as cathodic action when the current flows in the reverse direction if there be sufficient concentrations of alkali or lime salts in the surrounding electrolyte.

From the description of the occurrence and general characteristics of cable sheath corrosion in the present paper it may be concluded that although there are many conditions under which cables may corrode, the actual incidence of corrosion is small owing to the maintenance of non-corrosive chemical and electrical environments in the cable plant.

References

1. "Useful Information about Lead," Lead Ind. Assn. 1931, New York.
2. Archbutt, *Met. Ind.* (London), **18**, 341 (1921).
3. Gough and Sopwith, *Jour. Inst. Met.*, **56**, 55 (1935).
4. Burns, *Bell Sys. Tech. Jour.*, **15**, 20 (1936). For fuller details see Vernon, *Trans. Faraday Soc.*, **19**, 839 (1924); **23**, 113 (1927) from whose work the curves of Fig. 1 of this reference were adapted.
5. Pilling and Bedworth, *Jour. Inst. Met.*, **29**, 529 (1923).
6. Vernon, Discussion, *Jour. Inst. Met.*, **56**, 84 (1935).

7. Rawdon, *Bureau of Standards Scientific Paper 377* (1920).
8. Rawdon, *Indus. and Engg. Chem.*, **19**, 613 (1927).
9. Haigh and Jones, *Jour. Inst. Met.*, **43**, 271 (1930).
10. Haehnel, *Elekt. Nachrichten*, **2**, 74, 330 (1925); **3**, 97, 229 (1926); *Z. f. Metallk.*, **19**, 492 (1927). Lussaud and Noirclerc, *Rev. Gen. Elec.*, **34**, 19 (1933).
11. Townsend, *Proc. Am. Soc. Test. Mat.*, **27**, Part 2, 153 (1927).
12. Schumacher and Bouton, *Metals and Alloys*, **1**, 405 (1930).
13. Townsend, loc. cit.
14. Beckinsale and Waterhouse, *Jour. Inst. Met.*, **39**, 375 (1928).
15. Singleton, *Jour. Soc. Ind.*, **53**, 49T (1934).
16. Anderegg and Achatz, *Purdue Univ. Engg. Expt. Sta. Bull. 18* (1924).
17. Burns and Salley, *Indus. and Engg. Chem.*, **22**, 293 (1930).
18. Speller, Amer. Pet. Inst. "Drilling and Production Practice" (1935), p. 239.
19. *Bureau of Standards, Technical Paper No. 368.*
20. Logan and Taylor, *Bureau of Standards Journal of Research*, **12**, 119 (1934).
21. Haehnel, *Z. f. Fermeldetech. Werke u. Gerät.*, **4**, 35, 49 (1923).
22. Burns, loc. cit.
23. Bauer and Schikorr, *Metallwirtschaft*, **14**, 463 (1935).
24. Haring and Thomas, *Trans. Electrochem. Soc.*, **68** (1935). See also Burns, *Bell Sys. Tech. Jour.*, **15**, 37 (1936).
25. Jones, *Jour. Soc. Chem. Ind.*, **39**, 221T (1920).
26. Haehnel, *E.N.T.*, **8**, 77 (1931).
27. An experiment performed by Mr. G. D'Eustachio, Chief Chemist of the General Cable Corporation, and reported in a private communication.
28. Haehnel, *Z. f. Fermeldetech. Werke u. Gerät*, **4** (1923).
29. Evans, *Jour. Soc. Chem. Ind.*, **44**, 167T (1925).
30. Heap, *Jour. Soc. Chem. Ind.*, **32**, 771, 811, 847 (1913).
31. Höll, Gesundheits—*Ing.*, **58**, 323 (1935).
32. Shipley, *Jour. Soc. Chem. Ind.*, **41**, 311T (1922).
33. Whitman and Russell, *Indus. and Engg. Chem.*, **17**, 348 (1925).
34. Höll, loc. cit.
35. Friend, *Jour. Inst. Met.*, **39**, 111 (1928).
36. Friend and Tidmus, *Jour. Inst. Met.*, **31**, 177 (1924).
37. Müller, Reissmann and Ballin, *Z. phys. Chem.*, **114**, 129 (1924).
38. Bunker and Khachadoorian, *Jour. Worcester Poly. Inst.*, **26**, 12 (1922).
39. Brady, *Tech. Paper No. 8*, rev. ed. (1934), British Building, Research Board, Dept. Sci. & Ind. Res.
40. Blain, *Telephony*, **111**, July 4, 22 (1936).
41. Burns and Freed, *Jour. Amer. Inst. Elec. Engg.*, **47**, 576 (1928).
42. Burns and Clarke, *Indus. & Engg. Chem., Analytical Ed.*, **2**, 86 (1930).
43. Lehmann, *Z. F. Fermeldetech.*, **7**, 15 (1926); Garre, *Electrotechnische Zeit.*, **52**, 1418 (1931); DaFano, *Giorn. di. Chim. Ind. App.*, **14**, 18 (1932).
44. Thresh, *Analyst*, **47**, 500 (1922).
45. Britton and Evans, *Jour. Chem. Soc.* (London), 1773 (1930).
46. Friend and Tidmus, loc. cit.
47. Muir, Manchester Lit. & Phil. Soc. Proc., **15**, 35 (1875–6).
48. Haehnel and Klewe, *E.N.T.*, **9**, 407 (1932).
49. Haring, *U. S. Patent* 1,865,004, June 28, 1932.

Reduction of Airplane Noise and Vibration*

By C. J. SPAIN, D. P. LOYE and E. W. TEMPLIN†

THE three principal sources of airplane noise are the engine, the propeller, and air turbulence. Because of the impossibility of generating each kind of noise separately from the others, it has been necessary to develop what are in effect means for separating them and studying each one independently as they vary with speed of ship, speed of engine, and horsepower. In brief, the method that was used employs a series of tests under various flight conditions, the resulting data making it possible to solve a set of simultaneous equations. The paper gives numerous curves showing the variation with engine speed of the noise from these three sources.

Fundamental to any consideration of airplane noise are the characteristics of the ear itself. For the most part, physiology does not cooperate with the acoustical engineer when he sets out to increase the comfort of air travel. In fact, it has been necessary to develop several specialized measuring devices in addition to the familiar type of noise meter. Among these may be mentioned particularly a frequency analyzer which permits of selecting either a 20-cycle or a 200-cycle band out of any portion of the noise spectrum from 40 to 11,000 cycles per second. With the 200-cycle band filter the general shape of the noise characteristic is measured, while with the 20-cycle filter the frequency components of engine, propeller and other noise are identified and measured.

It is also desirable to be able to explore surfaces as to the extent to which they radiate noise. A microphone attachment has therefore been developed which quickly measures the characteristics of various interior surfaces. As a result, it has been found possible to improve the efficiency of distribution of the sound absorbing material, increasing its weight in certain locations and reducing it in others, thereby both lowering the noise level in the cabin and decreasing the total weight of acoustic treatment.

In order to measure the noise reduction provided by the cabin walls, another device known as a high-speed automatic level recorder has

* Digest of a paper presented before the Los Angeles (California) Section of the American Society of Mechanical Engineers, May 27, 1936. Published in full in *Transactions A.S.M.E.*, Vol. 58, No. 6, pp. 423–431, August 1936.

† The authors of this paper are engaged in technical work with the Electrical Research Products, Inc., a division of the Bell Telephone System.

been developed. It is in effect a rectifying and recording oscillograph whose stylus is capable of traveling at various speeds, the highest being such as to indicate in one second a difference of level of 840 db.

To reduce airplane noise within the passengers' and pilot's compartments, it is necessary to provide sound absorbing material as well as sound insulation. If there were no absorption within the cabin, the sound reduction would be zero, no matter how efficient the insulation, as the insulation would in this case only serve to delay the building up of the sound inside to the same intensity as outside. Laboratory equipment suitable for the study of absorbing materials and the measurement of their coefficients is therefore a very necessary adjunct.

Finally, mechanical vibration of audible rates to which various parts of a ship respond must be carefully studied. For this purpose, a so-called vibrometer has been perfected. With it, data are obtainable indicating the extent to which noise is transmitted into the cabin through the fuselage structure as compared to that coming through the air.

Abstracts of Technical Articles from Bell System Sources

The Renaissance of Physics.[1] KARL K. DARROW. Intended for the general public, this book is chiefly a story of some of the great discoveries and some of the grand general principles achieved or confirmed in physics since the century began. The title is an allusion to this period, for, to quote from the beginning of the book: "ever since the turn of the century physics has been enjoying a veritable renaissance, fairly to be likened with that splendid flowering of the arts and humane letters four hundred years ago to which the name of Renaissance was first applied. In this contemporary age when the artists in so many fields are overshadowed by the work of masters long since dead, the physicist has had the glorious good fortune of sharing in a spirit, an ambition, a sense of novelty and limitless opportunity, such as (we are told) inspired the Elizabethans."

The chapter headings run: *Physics and the Physicist—Intimations of Electricity—Release of Electrons from Matter—Through Measuring to Knowing—Magnets and Moving Charges—The Atom Visible—Light in the Semblance of Waves—Mystery of Waves and Corpuscles—Structure of the Atom—Technique of Transmutation—Victory over the Elements—Unity of Nature.*

There are forty-five illustrations, many of them half-tones of apparatus, spectra of various kinds, and processes of transmutation. No previous knowledge of physics is required of the reader, and the use of mathematics is confined to a few formulae of the simplest algebraical type. Much of the content of the book figured in the course of Lowell Lectures delivered by the author in Boston during the autumn of 1935.

Gutta-Percha—Effect of Vulcanization of its X-Ray Diagram.[2] C. S. FULLER. The finding of previous investigators that gutta-percha and balata have identical x-ray patterns is verified. Experiments on the x-ray behavior of vulcanized and unvulcanized gutta-percha show that vulcanization (to the extent carried out here) has no effect in changing the lattice plane spacings of either the alpha or beta crystal modifications. Vulcanization does appear to increase the degree of orientation of the crystallites present in these substances as produced by stretching

[1] Published by Macmillan Company, New York, N. Y., September, 1936.
[2] *Indus. and Engg. Chem.*, August, 1936.

and to that extent allows a more accurate calculation of the identity periods of the crystalline forms to be made.

A partial transformation of the beta to the alpha form of gutta-percha results by stretching at 80° C., although the exact conditions under which this occurs have not been determined.

The identity period in the fiber direction of the beta modification is 4.77 ± 0.03 Å., or double this value, and the alpha modification presents an anomaly in that two identity periods are in best agreement with the data. These are 9.00 ± 0.05 and 8.70 ± 0.13 Å. In the case of the beta modification three possible orthorhombic unit cells which are in agreement with the observed lattice plane spacings are given.

Fields Caused by Remote Thunderstorms.[3] K. E. Gould. The object of the studies described in this paper was to verify the supposition that certain types of short-duration longitudinal voltages appearing in communication circuits are caused by remote thunderstorms. By means of simultaneous directional measurements made in the frequency range below 40 kilocycles at two points as much as 900 miles apart, thunderstorms at distances of several hundred miles from one or both of these points have been located with a degree of accuracy great enough to permit conclusive correlation of the storm locations indicated by the directional measurements with the locations of recorded thunderstorms. Methods, equipment, and results are discussed.

Improved Types of Transmission Measuring Systems and Methods of Measurement.[4] W. H. Harden. The quantitative measurement of the electrical efficiency of telephone circuits as one of the important checks of the ability of these circuits to satisfactorily transmit speech has become an increasingly important maintenance function during the past twenty years. The function of transmission measuring equipment is not only to provide a convenient tool for quickly checking the electrical efficiency of telephone circuits, but also to serve as an aid in locating the cause of trouble when it is found to exist. It is the purpose of this paper to review briefly the progress which has been made in transmission testing technique and to describe some recent advances in the art which greatly facilitate this important part of telephone maintenance work. The discussion of these advances in the art will, we believe, be of interest to the railroads in connection with the operation and maintenance of their private telephone systems.

[3] *Elec. Engg.*, June, 1936.
[4] *Proc. Assoc. Amer. Railroads—Telegraph and Telephone Section*, June, 1935.

Some Improvements in Toll Circuit Design and Transmission.[5] GLEN IRELAND. Progress did not crash, along with the stock market, in 1929. Subsequent years have seen astonishing advances in many important businesses of the country as regards scientific developments, improved methods and better service. This is particularly true in the allied fields of transportation and communication where the service has been made more and more convenient, comfortable and accessible. Mr. Ireland's work lies in the field of communication and more specifically has to do with toll circuit design and transmission. In this paper he tells something of the progress in this field; first with respect to some new toll circuit instrumentalities that may be of direct interest in the work of the railroads, and secondly about some important and fundamental improvements, which are of general interest as indicating the trends in the art.

The general practices followed in connection with the design and installation of toll cables in the Bell System were described before the Telegraph and Telephone Section of the Association of American Railroads several years ago. There have been several specific changes in practices and some improved instrumentalities made available in this field which it is believed will be of interest to the railroads.

Calculated and Experimental Photoelectric Emission from Thin Films of Potassium.[6] HERBERT E. IVES and H. B. BRIGGS. Several years ago one of the writers proposed a theory of the photoelectric emission from thin films of alkali metals on supports of other metals, not photoelectrically active in the regions of the spectrum under observation. According to this theory the photoelectric emission is proportional to the rate of energy absorption by the thin film of alkali metal. The magnitude of the photoelectric current depends on the energy density immediately above the supporting metal, which is established from a knowledge of the optical constants of that metal, and upon the specific absorption of the alkali metal film. For its verification, the theory demands a knowledge of the optical constants of both supporting and alkali metals throughout the whole region of the spectrum where observations can be made. While optical constants have been determined for platinum, which is the metal most commonly used for a support for these thin films, no optical constants for the alkali metals have been available except in the visible region. In this region, a very satisfactory confirmation of the theory was obtained, particularly in respect to the variation of emission with the angle of incidence for the two principal planes of polarization (vectorial effect). One of the most characteristic phenomena of photoelectric

[5] *Proc. Assoc. Amer. Railroads—Telegraph and Telephone Section*, June, 1935.
[6] *Jour. Opt. Soc. Amer.*, June, 1936.

emission from thin films, namely, the occurrence of a pronounced maximum of emission in the spectrum, could not be compared with the predictions of this theory because these maxima in the case of the alkali metals lie in the ultra-violet. The theory of photoelectric emission from thin films has consequently had to stand unconfirmed in its entirety until such time as the optical constants of the alkali metals became available. In a separate paper the writers describe an experimental determination of the optical constants of potassium. In the present paper these constants are applied to the photoelectric theory, and the results are compared with experiment.

The Optical Constants of Potassium.[7] HERBERT E. IVES and H. B. BRIGGS. The importance of a knowledge of the optical constants of the alkali metals is emphasized by numerous recent theories of the metallic state and the optical properties of metals in general. In these theoretical treatments the alkali metals, because of their extraordinary properties, in particular their spectral region of transparency, have figured largely. There has, however, existed a serious gap in our experimental knowledge, in that optical constants have been entirely lacking for the region of extreme interest, namely, the ultra-violet. Without such knowledge theories must stand unchecked. Sufficient warrant for undertaking an experimental determination of the optical constants of the alkali metals, of which this study of potassium is the first, is therefore found on this ground alone. In addition, the writers have a special interest in these constants in connection with their work on the photoelectric effect. A theory of the photoelectric emission from thin films of alkali metals, proposed some years ago, which has been very successful in explaining the phenomena in the visible region of the spectrum, has urgently demanded optical data for its test in the ultra-violet region, where the most extreme and characteristic peculiarities of photoelectric emission are found.

Design and Equipment of a Fifty-Kilowatt Broadcast Station for WOR.[8] J. R. POPPELE, F. W. CUNNINGHAM, and A. W. KISHPAUGH. With its novel directional antenna, WOR produces a maximum field strength toward both New York and Philadelphia while limiting radiation in the direction of the ocean and sparsely populated areas. Radiation distribution measurements are given.

The layout of the station and the unique arrangements for lighting, heating, and ventilation of the building are described.

[7] *Jour. Opt. Soc. Amer.*, June, 1936.
[8] *Proc. I.R.E.*, August, 1936.

A serious attempt has been made to design and operate the equipment for a performance consistent with advanced ideas of high fidelity. Measurements from microphone to antenna of distortion, noise, and frequency response are presented.

Dial Switching of Connecticut Toll Calls.[9] W. F. Robb, G. M. McPhee, and A. M. Millard. The special application of step-by-step dial switching equipment to the handling of short distance toll telephone traffic was introduced in Connecticut in 1929, and has been extended gradually until at present approximately 46,000 toll messages per day, comprising 70 per cent of the traffic between exchanges in this area, are dispatched over the 1,367 circuits of the dial switching network. The resulting service improvements and savings in operating efforts are discussed in this paper, and a brief description of the transmission and equipment characteristics of the system is given.

[9] *Elec. Engg.*, July, 1936.

Contributors to this Issue

E. W. Bemis, Worcester Polytechnic Institute, 1919, B.S. in Electrical Engineering; Graduate Assistant in Electrical Engineering at Worcester Polytechnic Institute, 1919–1920. American Telephone and Telegraph Company, Department of Operation and Engineering, 1920–. Mr. Bemis has been engaged in work on carrier telephone systems, program transmission and public address systems, inductive coordination of power and telephone lines and since 1932 on telegraph engineering with particular reference to its transmission features.

R. M. Burns, A.B., University of Colorado, 1915; A.M., 1916; Ph.D., Princeton University, 1921; Instructor, University of Colorado, 1916–17. Western Electric Company, 1922–25. Bell Telephone Laboratories, 1925–; Assistant Chemical Director, 1931–. Dr. Burns' work has been largely in the electrochemical field and particularly on the subject of the corrosion of metals and its prevention.

James A. Carr, B.S. in Electrical Engineering, Virginia Polytechnic Institute, 1919; Graduate Student in Electrical Engineering at Massachusetts Institute of Technology, 1920. American Telephone and Telegraph Company, Development and Research Department, 1921–27; Bell Telephone Laboratories, 1927–. Mr. Carr's work has been mostly in aerial line design and maintenance studies.

B. L. Clarke, B.S., George Washington University, 1921; M.A., Columbia University, 1923; Ph.D., Columbia University, 1924. Bell Telephone Laboratories, 1927–. Dr. Clarke has been in charge of the work in analytical chemistry since 1930.

H. W. Hermance, Western Electric Company (Kearny), 1925–27, chemist; Bell Telephone Laboratories, 1927–. Mr. Hermance originally did general chemical analytical work. Since 1930 he has been in charge of the work in microanalysis.

A. D. Knowlton, B.S., Haverford College, 1920. Western Electric Company, Engineering Department, 1920–24. Bell Telephone Laboratories, 1924–. Mr. Knowlton is engaged in the development of telegraph equipment.

Frederick B. Llewellyn, M.E., Stevens Institute of Technology, 1922; Ph.D., Columbia University, 1928. Western Electric Company,

1923–25; Bell Telephone Laboratories, 1925–. Dr. Llewellyn has been engaged in the investigation of special problems connected with radio and vacuum tubes.

G. A. Locke, B.S., Cooper Union, 1920; E.E., Cooper Union, 1923. New York Telephone Company, 1908–15. Western Electric Company, Engineering Department, 1915–17. United States Signal Corps, 1917–19. Western Electric Company, Engineering Department, 1919–24. Bell Telephone Laboratories, 1924–. Mr. Locke is engaged in the development of telegraph switching circuits.

R. E. Pierce, Cornell University, 1913, A.B. and M.E. American Telephone and Telegraph Company, Engineering Department 1913–1919; Department of Operation and Engineering, 1919–. Mr. Pierce has been engaged in work on telegraph matters and during the period with the Department of Operation and Engineering has been in charge of telegraph engineering.

F. W. Reynolds, B.S. in Electrical Engineering, Union College, 1919; Ph.D., Cornell University, 1924; Signal Corps, U. S. A., 1918. American Telephone and Telegraph Company, Department of Development and Research, 1924–1934; Bell Telephone Laboratories, 1934–. Dr. Reynolds has been engaged in the development of telephotography.

F. J. Singer, B.S. in Electrical Engineering, University of Washington, 1920; M.S. in Electrical Engineering, University of Wisconsin, 1922. American Telephone and Telegraph Company, Department of Development and Research, 1922–34. Bell Telephone Laboratories, 1934–. As Telegraph Switching Engineer, Mr. Singer is engaged in the development of telegraph switching and testing facilities.